Abstracts of the Proceedings

iConference 2010

iSchools

FEBRUARY 3-6 • UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Program Committees

Conference Sponsors

We would like to thank the following companies and organizations for their generous support of the 2010 iConference:

GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE

YAHOO!

Microsoft Research

Microsoft is a registered trademark of Microsoft Corporation.

OFFICE OF THE VICE CHANCELLOR

Illinois Informatics Institute
Invent. Imagine. Innovate.

NSF

ISBN: 978-0-87845-128-9

Cover design: John Bonadies
Editor and Layout: Maeve Reilly
Proofreader: Leighton Christiansen

2010 ICONFERENCE PROGRAM CHAIRS:

Howard Rosenbaum, Indiana University, program co-chair

Karen Fisher, University of Washington, program co-chair

PLANNING COMMITTEE:

Alessandro Acquisti, Carnegie Mellon University

Harry Bruce, University of Washington

Elke Greifeneder, Humboldt-Universität zu Berlin

Joe Hall, UC Berkeley School of Information/Princeton University Center for Information Technology Policy

Susan Herring, Indiana University

Lori Kendall, University of Illinois at Urbana-Champaign

Xia Lin, Drexel University

Jeffrey Pomerantz, University of North Carolina at Chapel Hill

Jaime Snyder, Syracuse University

Mike Twidale, University of Illinois at Urbana-Champaign

Ping Wang, University of Maryland, College Park

Terry Weech, University of Illinois at Urbana-Champaign

Kate Williams, University of Illinois at Urbana-Champaign

LOCAL ORGANIZING COMMITTEE:

(from the University of Illinois at Urbana-Champaign, unless noted)

Cindy Ashwill

Carie Burgess, Microsoft

Matt Beth

Heekyung Choi

Ingbert Floyd

Sean Goggins, Drexel University

Jonathan Grudin, Microsoft

Sharon Johnson

Natalie Michael, University of Washington

Maeve Reilly, iSchools

Allen Renear

Kim Schmidt

John Unsworth, chair

Richard Urban

Dan Wright

STUDENT VOLUNTEERS:

Students from the Graduate School of Library and Information Science at the University of Illinois

ADDITIONAL THANKS TO:

iSchool Communicators

From the Program Committee Chairs

Welcome to the Fifth Annual iConference, this year at the University of Illinois at Urbana-Champaign. The 2010 iConference theme addresses the impacts of the iSchool movement inside and outside our community. More than 300 scholars, professionals, and students from diverse backgrounds are gathering for four days to explore, discuss, and debate the complex issues at the nexus of people, information, and technology.

In 40 sessions (20 paper sessions, 14 roundtables and 6 wildcard sessions), we will consider two overarching questions:

- What are the broad impacts (actualized and potential) of the iSchool movement?
- How can impact be defined, identified, measured, and communicated to key audiences?

We will also reflect upon the core activities of the iSchool community, including research, design, methods and epistemologies, educational practices, and engagement between the iSchools and wider constituencies both in the United States and abroad. Contributions also reflect more broadly on complex interrelationships among people, information, and technology in the iSociety, particularly those focusing on public and private sector settings.

Overall there were 239 submissions to this conference. We are pleased that 53 papers, 14 roundtables, 6 wildcards and 78 posters have been accepted. For your convenience, there are seven tracks into which the papers have been placed: diversity, information retrieval, information organization, information behavior, e-government and e-science, iSchool curriculum and pedagogy, and a special iSchool associate deans' research track. The posters are organized into eight categories: diversity, digital libraries, e-government and e-science, information behavior, information management, information organization, information retrieval, and the iSchools.

With its plenaries, paper and poster sessions, roundtables, wildcard sessions, preconference workshops, doctoral colloquium, mentoring sessions, and ample opportunities for conversations and connections at the reception, lunches, breaks, banquet, and evening discussions, the iConference celebrates and engages our multidisciplinary and diverse research communities, drawing on the interest and expertise of people across disciplinary and organizational boundaries. We invite you to engage with your iSchool colleagues throughout the conference, learn about each others' work and consider the many ways that we create impact.

Howard Rosenbaum
Associate Dean, School of Library and Information Science, Indiana University

Karen E. Fisher
Professor, The Information School of the University of Washington, and Chair of the Information & Society Center (ISC)

Reviewers

2010 ICONFERENCE REVIEWERS:

Donato Barbagallo
Karine Barzilai-Nahon
Randolph G. Bias
Catherine Blake
Leanne Bowler
Kathy Burnett
Carlos Caicedo
Kelly Caine
Hong Cui
John M. Daughtry
M. Carl Drott
Megan Finn
Andrew Fiore
Karen Fisher
Renee Franklin
Snehal Gaikwad
Cheong Hian Goh
Elke Greifeneder
Joe Grobelny
Joseph Hall
Stephanie Haas
Derek Hansen
Daqing He
Shaoyi He
Libby Hemphill
David Hendry
Susan Herring
Chris Hinnant
Shuyuan Mary Ho
Xiao Hu
Lori Kendall
Gerald Knezek
Susan Knoer
Kyungwon Koh
Saranga Komanduri
Cathy Kozlowicz
Xia Lin
Jingjing Liu
Ying-Hsang Liu
Kelly Lyons
Marta L. Magnuson
Carleen Maitland
W. John MacMullen
Bettie McGinness
Helen Mentis
Eric T. Meyer
Shawne Miksa
Carlos Monroy
Kim Nimon
Brian C. O'Connor
Mark Pfaff
Jeffrey Pomerantz
Devendra Potnis
Howard Rosenbaum
Terrell Russell
T-Kay Sangwand
Steve Sawyer
Pnina Shachaf
Linda Smith
Jaime Snyder
Andrey Soares
Mega M. Subramaniam
Lindsay Tabas
Mike Twidale
Vetle Torvik
Rajesh Veeraraghavan
S. Vijayalakshmi
Nina Wacholder
Ping Wang
Scott Warren
Martin Weiss
Mitch Wiesberg
Kate Williams
Megan Winget
Dinghao Wu
Sarita Yardi
David Yates
Bei Yu
Michel Zimmer

Table of Contents

Papers

iConference 2010

FEBRUARY 3-6 • UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

What to Do with a Million Pages of Digitized Historical Newspapers?

Robert B. Allen
College of Info Sci & Tech
Drexel University
Philadelphia, PA, 19107
1-215-895-0460
rba@drexel.edu

Weizhong Zhu
College of Info Sci & Tech
Drexel University
Philadelphia, PA, 19107
1-215-668-4185
wz32@drexel.edu

Robert Sieczkiewicz
Hagerty Library and Drexel Archives
Drexel University
Philadelphia, PA
1-215- 895-1757
robs@drexel.edu

ABSTRACT
Newspapers are rich sources of evidence of history and literally millions of pages of historical newspapers have now been digitized. We aim to develop tools for effectively browsing this rich resource. However, the structure of newspapers is highly complex and a complete analysis will involve many interacting components. We demonstrate two approaches for extracting advertisements from other types of material in the newspaper. We also describe preliminary results of interviews with historians about features which they would find particularly useful for conducting research on collections of digitized historical newspapers.

Categories and Subject Descriptors
H.3.3 Information Search and Retrieval; H.3.7 Digital Libraries; H.5.2 User Interfaces

General Terms
Design, Human Factors

Keywords
Advertisements, History, Image Processing, Interface Design, Newspapers

1. DIGITAL HISTORY AND DIGITIZED HISTORICAL NEWSPAPERS
Along with several other areas of the humanities, information systems are greatly expanding the resources available at the desktop for historians. Literally millions of pages of historical newspapers have already been digitized and many millions more will be processed over the next few years.

While there are several digitized initiatives underway, we have worked with the files prepared by the NDNP (National Digital Newspaper Program). This is a joint project of LC and NEH which builds on the prior USNP (United States Newspaper Program). USNP had worked with state libraries to prepare archival quality microfilm of historical newspapers from their sates and that microfilm is now being digitized by NDNP. A notable aspect of the NDNP digitization is that public domain OCR files are delivered in the METS-ALTO format. In almost all other newspaper digitized projects the OCR is proprietary or otherwise unavailable. However while the OCR is available in NDNP, it is often of marginal quality and extensive text processing is necessary.

Figure 1: Sample page of the *Washington Times* for March 1904.

2. TEXT PROCESSING FOR EXTRACTING ADVERTISEMENTS
There are many approaches for text processing of the historical newspaper text. [2]. We have demonstrated extraction and categorization of articles is [3]. While those results are uneven, there are many additional constraints which can be considered. One technique which was explored by Allen and Hall [4] was to focus the first lines of text in a segment to find items which were repeated across days. This was used to identify features stories but it was also noted that after extracting the longer feature articles that many of the shorter items with repeated headings were advertisements. This is sensible since advertisements often run for many days while other items change from day to day.

Here, we study this observation in more detail. Specifically, we applied the article segmentation procedure described in [2] to OCR output for the *Washington Times* for March 1904. We then processed that to find repeated first lines of the text

segments. Though, these had to be heavily processed because of the large number of OCR errors.

J WILLIAM LEE

E M EARLE SON

THE STORE THAT SAVES YOU MONEY

NATIONAL BISCUIT COMPANY

ADVANCE SPRING STYLES

Table 1: Examples of repeated headings indicating advertisements.

We found three categories among the repeated items: feature articles, headings for advertisements, and headings for other special sections such as Vital Records. Most of the advertisements appeared on the right and bottom edges of the newspaper (see Figure 1). Moreover, they all had non-standard fonts which probably could have been helpful for identifying them.

While this technique does not capture all of the advertisements, many of them are clustered there. Furthermore, we can infer that other advertisements are located in the same region even if they don't have repeated first lines. However, as we noted in our earlier work [3, 4], there needs to be a design tradeoff between the complexity of automated inference and the amount of human knowledge to include but the simple rules of thumb described here caught the large majority of advertisements.

3. AUTOMATIC ANALYSIS OF IMAGE GENRES

As a second, very different type of analysis for detecting advertisements is to determine image genres[1]. Images with words are most often banners for advertisements on the other hand, portraits may accompany news stories. In addition to the news codes for text used by [3], the International Press and Telecommunication Council (IPTC) also specifies image genre types[2]. A complete system for processing newspaper images would itself be a major project. The goal here is not to extract the images from the page but to categorize then by the IPTC genre codes after they have been extracted.

3.1 Image Feature Selection and Representation

There are several approaches to image analysis. One common strategy looks for specific features. We applied Eigenvector-based feature selection [5] and traditional retrieval/classification methods to differentiate images from the different genres. Models of individual features try to identify position, size, color, texture and relationships between these features but these models are insufficient to classify the multiple views such as multiple faces in one image, compared to the models of gestalt/pattern features. Because the task includes complex images such as images with multiple persons, Eigenvector based feature modeling seems to be more suitable for our project. The Eigenvector based feature indexing was originally developed for face recognition and includes the following steps:

- Take the M training image vectors and average them to find $\Psi = 1/M * \sum_{n = 1 \text{ to } M} \Gamma_n$ where Γ_n is the n^{th} image vector. Each image differs from the average by $\Phi_i = \Gamma_i - \Psi$.
- These are the eigenvectors of the Covariance Matrix: $C = 1/M * \sum_{n = 1 \text{ to } M} \Phi_n \Phi_n^T = AA^T$, where $A = [\Phi_1 \ldots \Phi_M]$. If $M << N^2$ then there will only be $M - 1$ eigenvectors which have non-zero Eigenvalues. So can solve an M x M matrix instead. Consider the eigenvectors, $\mathbf{v}_i$, of A^TA such that: $A^TAv_i = \mu_i v_i$ which yields $AA^TAv_i = \mu_i Av_i$ where Av_i are the eigenvectors of $C = AA^T$ (and μ_i are the Eigenvalues).
- To form the Eigenspaces u_i use the following equation: $u_i = \sum_{k = 1 \text{ to } M} v_{ik}\Phi_k$ for i = 1 ... M.

Each image in the Eigen indexing spaces is represented as a vector with at most M dimensions. If a new image is projected in the training Eigenspaces, the representation of that image is calculated by the following procedure:

- Given the set of M Eigenspaces, choose the M ' Eigenspaces that have the highest associated Eigenvalues.
- Take a new image, Γ, and project it into "Eigenspaces" by the operation: $W_k = u_k^T \cdot (\Gamma - \Psi)$ for k = 1 to M '.
- The weights (W_k) form a vector $\Omega^T = [W_1 \ldots W_k]$ which describes the contribution of each Eigen-space in representing the input face image.

3.2 Image Retrieval and Classification Methodology

To match the query image, the relevance between the query image and each of the training samples is defined by the Euclidian distance $\varepsilon_k = \|\Omega - \Omega_k\|$, where Ω_k is the weight vector describing the k^{th} training image. But to classify an image, the representation of the image is the vector Ω^T. The traditional classification methods include two types, unsupervised learning – clustering and supervised learning – machine learning. Two clustering algorithms: basic K-means [6] and Kohonen Mapping [7] and two machine-learning techniques, back-propagation [8] and simulated annealing [9] were tested.

3.3 Test Corpus of Images

A preliminary examination of the newspapers found that six of the IPTC image genres covered most of the cases. A research assistant obtained 60 newspaper images from the *Washington Times* for 1904 belonging to these six classes: Portrait (PRT), Text Characters (WORD), Exterior Views of Buildings (EV), Full Body (FB), Half Body (HL), and a Group of People (MP). There were more of the first three types than of the others. Examples of the images are shown in Figure 2. They were normalized to a fixed size before the analysis was conducted.

[1] The entire page scanned from the microfilm is also an image but here we are concerned with images extracted from within the page.

[2] http://www.iptc.org/cms/site/index.html;jsessionid=aiZ0zp rArDm8?channel=CH0089

	Kohonen	K-means
Purity	*0.55*	*0.42*
Entropy	*0.50*	*0.68*
NMI	*0.47*	*0.31*

Table 2: Comparison of clustering methods.

3.4 Clustering the Images

Two clustering techniques, Kohonen Mapping and basic K-means, are used to automatically classify the six categories of the images. In Table 2, three evaluation metrics, Purity, Entropy and normalized mutual information (NMI) are used to evaluate the performance of the two techniques. The values of the three measures range from 0.0 to 1.0. The higher the values of Purity and NMI are, the better the performance of the method is. The lower the value of Entropy is, the better the performance of the method is. The results in Table 3 indicate that for this dataset, Kohonen mapping out-performs basic K-means for every evaluation method.

Figure 2**: Examples of the six image genre classes (upper row: HL, WORD, EV, lower row: PTR, HL, MP)**

Clusters generated by the Kohonen Mapping are shown in 3. There are strong associations: Cluster 1 and PRT; Cluster 2 with MP; Cluster 3 with FB; Cluster 4 with ENV; and Cluster 5 with WORD; but Cluster 6 did not have a clear association and it might be deleted. Similarly, the half-body images might be merged with the full-body images.

3.5. Machine Learning of the Genre Categories

Two neural network-based machine learning methods, Back-propagation and Simulated Annealing were used to automatically classify the six categories of the images. Eighty percent of samples were randomly picked for training and the remaining 20% used for testing 10 or 20 times to calculate an average learning performance. In Table 4, PRT/HL/FB/MP/WORD/EV denotes that the training set includes six classes; PRT/HL/FB/MP denotes the four PERSON related classes; PERSON/WORD/EV denotes the three major categories, PERSON (PRT/HL/FB/MP), WORD and EV; PRT/WORD denotes the two classes, PRT and WORD. The results indicate that on this dataset, Back-propagation out-performs Simulated Annealing for every type of training set. Further, the two machine learning methods did better on the two training sets, PERSON/WORD/EV and PRT/WORD.

	Portrait	WORD	Environment	Full Body	Half Body	Multiple People
Cluster 1	*9	1	1	0	1	0
Cluster 2	0	0	1	0	0	*4
Cluster 3	4	0	1	*7	2	*4
Cluster 4	1	0	*6	1	2	0
Cluster 5	2	*9	0	0	1	0
Cluster 6	0	0	0	1	0	1

Table 3: Number of each type of image in the clusters. * indicates the best match.

Training Sets	Back-Propagation	Simulated Annealing
PRT/HL/FB/MP/WORD/EV	0.46	0.23
PRT/HL/FB/MP (Person)	0.33	0.27
PERSON/WORD/EV	0.64	0.40
PRT/WORD	0.75	0.63

Table 4: Comparison between Back-propagation and Simulated Annealing for categorization.

3.6. Image Retrieval with Query-by-Example

Image retrieval may use "query-by-example" compared to text retrieval. Our approach identifies the most relevant or similar images in the Eigen-indexing spaces to the query image by projecting it to the indexing spaces. In Table 5 one PRT image and one WOPRD image were randomly selected and then treated as queries and the top five most relevant images in the indexing spaces are listed. As can be seen, the query-by-example returns a high number of images from the same genre class.

WORD Image	***Similarity***
WORD_March_19_1904_Page9.jpg	*target*
FB_Jan_3_1904_Page3.jpg	*0.72*
WORD_March_13_1904_Page10.jpg	*0.63*
WORD_Jan_3_1904_Page12.jpg	*0.62*
WORD_March_18_1904_Page12.jpg	*0.58*
HL_Jan_18_1904_Page7.jpg	*0.57*
PRT Image	***Similarity***
PRT_March_3_1904_Page3.jpg	*target*
PRT_Jan__1904_Page2.jpg	*0.76*
PRT_Untitled.jpg	*0.76*
WORD_March_21_1904_Page5.jpg	*0.70*
MP_Jan_3_1904_Page8.jpg	*0.66*

Table 5: Two cases of query-by-example retrieval for images.

3.7. Image Genre Discussion and Conclusion

The clustering and machine learning methods categorize the three major categories, WORD/PERSON/VIEW effectively with Eigenspace indexing but differentiate the sub-categories of PERSON poorly. This is a small study, but it suggests that the image modeling of Eigenspace based features are good on image categorization with fewer classes, for instance, binary image categorization. For this data set, Kohonen mapping is a better clustering method than basic K-means, and Back-propagation is a better machine learning method than Simulated Annealing. In the future, a much larger collection will be tested for the proposed approach and integration with specific individual features and textual descriptions will be explored to improve the performance of the Eigenspace indexing model and automatic categorization.

4. INTERFACE REQUIREMENTS FOR SUPPORTING HISTORIAN'S ACCESS TO DIGITIZED NEWSPAPERS

There have been quite a few studies of the information needs of historians [10-13]. However, these studies do not provide clear guidance about what sort of interface tools historians would find most useful for interfaces with a collection of newspaper. The primary interface for searching the NDNP newspaper collection is Chronicling America[3] which is a basic search interface with few features.

There have been some ad hoc recommendations such as supporting searching for Facts, Trends, Searching for Details[4]. Tools could be based on the proposals of "Rachel"[5]

(1) Have a List,
(2) Find a thread of some kind,
(3) Don't just use one newspaper,
(4) Don't fall into the trap of only reading articles that your keywords throw up.

[3] http://chroniclingamerica.loc.gov/

[4] Rubenstein, A., Center for History and New Media, Unpacking Evidence, http://chnm.gmu.edu/worldhistorysources/unpacking/newshow.html (accessed November 2009)

[5] Rachel, A Historian's Craft: http://idlethink.wordpress.com/2009/06/16/on-newspapers-as-sources/ (accessed November 2009)

(5) Use existing secondary literature,
(6) Keep really, really scrupulous notes, and
(7) Don't neglect the letters and the advertisements.

This list suggests a number of potential features.

To explore the importance of these services by working historians in more detail, we conducted informal interviews with two historians. Here are some examples of notes from these interviews:

> The *Chicago Tribune* database is good for searching names, but broader topics are hard to research – e.g., race relations brings back too many results.
>
> A log of all searches – `this is a huge issue for me'. Editing a book manuscript recently, she found it `hugely taxing' to find items she hadn't cited.
>
> Searches lead to other searches, so she would like ways to see how searches are nested within each other and to get back to earlier search results. A visual map telling you where you are in your search would be especially helpful. A system that lets her easily use multiple windows.
>
> [The historian] used newspapers to fill in gaps in research and corroborate information from other sources. Exploratory searching included looking at larger issues and events such as elections and campaigns. She used newspapers to find public opinion about changes in liquor license laws – to get a sense of `the texture of the city… how the city was thinking'.

It is clear that the historians will benefit from an interface which would support richer ways of interacting with the collection than are currently supported by the Chronicling America web site. We are collecting more interviews and using those comments for an interface prototyping effort.

5. FUTURE OF ACCESS TO HISTORICAL NEWSPAPERS

There are a great many challenges to supporting effective access for the newspapers in NDNP collection. Yet, that is just the beginning. It will have to be scaled up to several fold to incorporate newspapers from around the world. Moreover, that can be scaled that much more again, to support links to other historical resources such as books, images, and manuscripts.

We are exploring techniques for coordinating among historical resources. As a first step, multiple newspapers from one town or region can show synergies which improve the text processing of each. In the same vein, other historical resources such as biographical and building databases can be cross referenced with the newspapers. Finally, we also envision novel tools for describing and interacting with "threads of history" and we are actively developing formalisms for describing chains and historical events and timeline interfaces for visualizing them (e.g., [14]).

6. ACKNOWLEDGMENTS

We thank faculty from the Drexel University Department of History & Politics for interviews. We also thank Mike Zarro for assistance. Weizhong Zhu completed his doctorate and is now employed at WebLib Inc.

7. REFERENCES

[1] Schudson, M., (1978). *Discovering the News: A Social History of American Newspapers.* Basic Books, New York.

[2] Allen, R.B., Japzon, A., Achananuparp,P., and Lee, K-J., (2007). A Framework for Text Processing and Supporting Access to Collections of Digitized Historical Newspapers. *HCI International Conference*.

[3] Allen, R.B. and Hall, C., (in preparation) Automated Processing of Digitized Historical Newspapers beyond the Article Level: Finding Sections and Regular Features.

[4] Allen, R.B., Waldstein, I., and Zhu, W., (2008). Automated Processing of Digitized Historical Newspapers: Identification of Segments and Genres. *ICADL*, 380-387.

[5] Turk, M.A. and Pentland, A.P., (1991) Face recognition using Eigenfaces," (1991) *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 586-591.

[6] Dubes, R.C., and Jain, A.K. (1988). *Algorithms for Clustering Data*, Prentice Hall, New York.

[7] Kohonen, T. (1990). The Self-organizing Map. *Proceedings of the IEEE 78*(9), 1464–1480.

[8] Rojas, R., "The Backpropagation Algorithm", (1996) *Neural Networks – A Systematic Introduction*, Springer-Verlag, Berlin, New York, ISBN 9783540605058.

[9] Kirkpatrick S., Gelatt C. D., and Vecchi M. P., (1983) "Optimization by simulated annealing", *Science*, vol. 220, 671–680.

[10] Tibbo, H.R., (2002). Primarily History: Historians and the search for primary source materials. *ACM/IEEE Joint Conference on Digital Libraries*, 1-10.

[11] Bradshaw, J. (1984). The Use of Newspapers in Historical Research, *Chronicle* (East Lansing) (04409426), 20(2), 18-19.

[12] Case, D. (1991). The Collection and Use of Information by Some American Historians: A Study of Motives and Methods. *Library Quarterly*, 61(1), 61-82.

[13] Dalton, M., and Charnigo, L. (2004). Historians and their Use of Information Sources. *College & Research Libraries*, 65(5), 400-25.

[14] Allen, R.B. (2005). A Focus-Context Timeline for Browsing Historical Newspapers. *ACM/IEEE Joint Conference on Digital Libraries*, 260-261.

Measuring Research Impact: A First Approximation of the Achievements of the iSchools in ISI's Information and Library Science Category – An Exploratory Study

Judit Bar-Ilan
Department of Information Science, Bar-Ilan University
Ramat Gan, 52900, Israel
barilaj@mail.biu.ac.il

ABSTRACT
In this paper, we analyze those publications of the home institutes of the iSchools that are indexed by Thomson Reuters (ISI) Web of Science in the information science and library science category, and were published between 2000 and 2009.

Categories and Subject Descriptors
Could not find an appropriate ACM category.

General Terms
Measurement.

Keywords
Research evaluation, information and library science, publications, citations, h-index

1. INTRODUCTION
This year the theme of the iConference is iMPACTS, including the research impact of the iSchools [12]. Thus we decided to assess the research impact of the iSchools on information and library science, by retrieving all the items indexed by Thomson Reuters (ISI) Web of Science in the subject category "information and library science" that were published by the iSchools' home institutions in the period 2000-2009. We measured the number of publications, the number of citations, the h-index of the set of retrieved items, the most highly cited item, the most frequently appearing document type and the journal in which the highest number of items were published by the specific institution during the whole period.

This method has limitations it can only approximate the iSchools' research performance. It is quite obvious that some of the iSchools are not very active in the subject area defined by ISI as "information and library science". These schools are probably more active in other areas like computer science or information systems. On the other hand, it is possible that some of the publications of the given university in the subject category "information science and library science" were not produced by members of the iSchool, but rather by members of other departments who publish in journal in the "information science and library science" category. In addition there are indexing mistakes, and sometimes the affiliation of the author does not appear on the paper, and thus are not counted for the given institution. Nonetheless, the results of this exercise can serve as a first estimate.

Early ranking studies were based on perceptions and rankings were provided by survey participants (e.g. [26, 27]). Mulvaney [17] analyzed White's finding and found that perceived quality is associated to some extent with faculty productivity, but this was not the most influential variable. On the other hand Biggs and Bookstein [3] interviewed 45 randomly selected faculty members from ALA accredited schools and asked them what constitutes a high quality MLS program. The only criterion which was mentioned by the majority of the respondents was the presence of faculty members who are active in research and publishing.

Danton [10] reviewed and compared eight early rankings of library and information science schools, some of the rankings were based on perceptions, while others on citation and publication counts. Cronin and Overfelt [9] strongly questioned the perception studies, and conducted a very thorough analysis of a single information and library science school. They reached the conclusion that publication and citation counts are heavily skewed by a few bibliometric stars, supporting the conclusions of Brace [5].

Besides perception based rankings, there were studies based on publication counts only. Boyce and Hendren [4] based their study on data retrieved from Library Literature. They counted publications with and without book reviews. They also normalized the numbers by the number of full time faculty in each of the institutions. Wallace [25] also based his study on data from the Library Literature databases. Varlejs and Dairymple [24] retrieved data from Library Literature, LISA and ISA; while Pettigrew and Nicholls [20] conducted a large study using data from ERIC, LISA, PASCAL, LLIT and SSCI. They reached the conclusion that schools offering PhD programs are more productive than schools without such programs. Meho and Spurgin [15] reviewed several of the earlier studies that aimed to rank library and information science schools based on research productivity. They retrieved data from nine databases to achieve

reasonable coverage of the publications of the top 20 individuals listed in Budd's study.

Quite a number of studies took into account in addition to publication counts citation counts as well. One of the earliest studies of this type was conducted by Hayes [13], who used the Social Science Citation Index (SSCI) as his data source, and covered the years 1965-1980. Budd and Seavey [7] carried out a follow-up study covering the years 1981-1992. They provided publication and citation counts and also normalized data per capita. Budd [6] and Adkins and Budd [1] conducted additional follow-up studies covering the years 1993-1998 and 1999-2004 respectively.

Bates [2] emphasized the need to take into account books when evaluating LIS faculty. Note that books are not indexed by the Web of Science (WOS) and thus are not taken into account in the current study. She combined three types of data: perceptions of quality, publication and citation counts, differentiating between different types of publications: articles, books, edited books or journal issues and book reviews, letters and editorials.

Shaw and Vaughan [23] profiled a "typical" LIS professor, based on his/her publication and citation patterns. Meho and Yang [16] used the Web of Science, Scopus and Google Scholar to rank LIS researchers based on citation counts retrieved from the three citation databases. Seng and Willet [22] examined the citedness of UK library school publications. The citation counts correlated highly with the results of the RAE exercise. Such correlation was also found by Oppenheim [18]

With the introduction of the h-index [14], a number of studies were carried out ranking information scientists in the US and in the UK [8, 19, 21] according to their h-indices.

2. METHODS

The iSchools movement currently has 27 members. For operational reasons we had to combine the outputs of the two Indiana University at Bloomington iSchools, the School of Library and Information Science and the School of Informatics. The reason for this was that we used the OG field tag (organization) of the Web of Science, and did not want to use the SG, the suborganization field tag so as not to decrease the recall (in case authors did not specify the name of their school). For the Indiana University we also added Bloomington to the search, to exclude the publications of the School of Library and Information Science in Indianapolis.

Using the name of the university only, and limiting the search to the subject category "information science and library science" obviously excluded some of the publications of the iSchools and obviously included other publications of the university where the authors were not members of the iSchool. As stated above, this study is only a first approximation of the outputs of the iSchools in the area of information and library science. The publication years were limited to 2000-2009.

For each iSchools we calculated the number of publications, the number of citations, the h and g-index of these publications, the most frequently occurring publication type, and the frequency of the publication type. In addition, we tabulated the most highly cited publication of each iSchool, the number of citations this item received, the most productive faculty member in the category "information science and library science", and the number of his/her publications indexed and the journal in which the iSchool published the largest number of publications.

The h-index for a set of citable items of size N is defined following Hirsch [14] who defined the h-index for authors, as the unique number h, such that h items are cited h times or more, and the remaining N-h items are cited h times or less.

One of the shortcomings of the h-index is that it does not take into account the access citations of the top-cited items. Thus Egghe [11] introduced the g-index. A set of N items has g-index g, if these items are ordered in decreasing order of the number of citations they received and g is the highest rank, such that the top g items received at least g^2 citations.

The data was collected from the Web of Science's Social Science Citation Index, but without the Conference Proceedings Citation Index for the Social Sciences. Thus only proceedings papers indexed by the Social Science Citation Index were included. Data was collected between November 2 and 5, 2009.

3. RESULTS AND DISCUSSION

In Table 1 the publication and citation counts of the home institutions of each of the iSchools (the two Bloomington iSchools have the same home institution) in the subject category "information science and library science" for the year 2000 and 2009 are displayed. In addition the table shows the h and g-indexes and the most frequently occurring publication types.

The results show that the largest number of publications was produced by the University of Illinois, and the largest number of citations was accumulated by the University of Maryland. For a few of the universities the publication and citation counts are extremely low, indicating that even though they have an iSchool they do not publish heavily in the areas of information and library science. They are probably active in other areas relevant to iSchools like computer science and/or information systems.

In terms of the h and g-indexes, the highest numbers were achieved by Indiana University and the University of Maryland respectively. Note the huge difference between the highest h-index (27) and the highest g-index (56). The h-index of the University of Maryland papers was only 21, but because this university's top-cited papers were cited many more than 21 times, the g-index reached the value of 56.

With a few exceptions, the most frequently occurring publication type was "article". Some universities (Drexel, Georgia Tech, Illinois, Penn State and Rutgers) published more book reviews than articles. For Georgia Tech, 66.5% of the publications were book reviews.

Table 1: Publication and citation counts, h and g-index and major publication type of the host institutions in the subject catgory "information science and library science" during the period 2000-2009

University	Abbrev.	Publication count	Citation count	h-index	g-index	Most frequent publication type	Count of most frequent type
University of California, Berkeley	UCBer	102	353	11	18	article	54
University of California, Irvine	UCIrv	70	734	13	26	article	42
University of California, Los Angeles	UCLA	304	1054	16	28	article	112
Carnegie Mellon University	CMU	98	476	11	18	article	58
Drexel University	DRXL	269	1518	18	37	book review	128
Florida State University	FLST	271	1279	15	31	article	140
Georgia Institute of Technology	GTEC	158	207	8	12	book review	105
Humboldt-Universität zu Berlin	HUMB	51	37	5	5	article	28
University of Illinois	ILUC	920	1571	20	28	book review	515
Indiana University, Bloomington	INDB	451	2484	27	41	article	179
University of Maryland	UMLND	388	3689	21	56	article	184
UMBC	UMBC	9	20	3	4	article	5
University of Michigan	Umich	254	1229	18	28	article	130
University of North Carolina	UNC	408	1878	21	36	article	208
University of North Texas	UNT	80	186	7	12	article	37
The Pennsylvania State University	Penn	588	2194	21	39	book review	287
University of Pittsburgh	Pitt	354	1861	22	35	article	150
Royal School of Library and Information Science, Denmark	Denmk	145	980	15	26	article	78
Rutgers, the State University of New Jersey	Rutg	388	1630	19	37	book review	190
University of Sheffield, England	Sheff	206	875	17	22	article	146
Singapore Management University	Sing	7	32	2	5	article	6
Syracuse University	Syrac	190	665	12	20	article	98
University of Texas, Austin	UTA	399	1586	21	31	article	208
University of Toronto	UTor	253	627	12	21	article	104
University of Washington	UW	485	1286	18	30	article	194
Wuhan University, China	Wuhan	43	74	4	7	article	40

Table 2 displays the most highly cited publication of each institution, the number of citations it received and the publication type.

The most highly cited item is by Venkaesh et al. Although Venkatesh was at the time of writing affiliated with the University of Maryland, but he was not at the College of Information Studies, but at the School of Business. Currently he is affiliated with the University of Arkansas.

Among the most highly cited papers, seven of them were published in the MIS Quarterly and five of them in the Journal of American Medical Informatics Association (JAMIA). Neither of these journals are considered to be core information or library science journal. This a well-known problem with the ISI category of information and library science.

The top cited item is responsible on average for 16.40% of the total citations received by the university, the percentage ranges between 87.50% (University of Singapore) and 4.48% (University of Texas at Austin).

Table 2. Most cited publications

Univ.	authors	abbreviated title	source	publ year	TC	publ type
UCBer	Patel, VL; Arocha, JF; Kaufman, DR	Review? A primer on aspects of cognition for medical informatics	JAMIA	2001	40	review
UCIrv	Melville, N; Kraemer, K; Gurbaxani, V	Review: Information technology and organizational performance	MIS QUART	2004	129	review
UCLA	Bates, DW; Cohen, M; Leape, LL; Overhage, JM; Shabot, MM; Sheridan, T	White paper - Reducing the frequency of errors in medicine using information technology	JAMIA	2001	191	article
CMU	Chen, PY; Hitt, LM	Measuring switching costs and the determinants of customer retention in Internet-enabled businesses	INF SYST RES	2002	73	article
DRXL	Gefen, D; Karahanna, E; Straub, DW	Trust and TAM in online shopping	MIS QUART	2003	348	review
FLST	McKnight, DH; Choudhury, V; Kacmar, C	Developing and validating trust measures for e-commerce	INF SYST RES	2002	225	review
GTEC	Dietz, JS; Chompalov, I; Bozeman, B; Lane, EO; Park, J	Using the curriculum vita to study the career paths of scientists and engineers	SCIENTOMETRICS	2000	23	article
HUMB	Fritsche, L; Schlaefer, A; Budde, K; Schroeter, K; Neumayer, HH	Recognition of critical situations from time series of laboratory results by case-based reasoning	JAMIA	2002	7	article
ILUC	Ranganathan, C; Ganapathy, S	Key dimensions of business-to-consumer web sites	INF& MAN	2002	103	article
INDB	Kling, R; McKim, G	Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication	JAMIA	2000	110	article
UMLND	Venkatesh, V; Morris, MG; Davis, GB; Davis, FD	User acceptance of information technology	MIS QUART	2003	651	article
UMBC	Rubenstein-Montano, B; Buchwalter, J; Liebowitz, J	Knowledge management: A US Social Security Administration case study	GOV INF QUART	2001	5	article
Umich	Saha, S; Saint, S; Christakis, DA	Impact factor: a valid measure of journal quality?	J MED LIBR ASS	2003	89	article
UNC	Gold, AH; Malhotra, A; Segars, AH	Knowledge management: An organizational capabilities perspective	J MAN INF SYST	2001	171	review
UNT	Beatty, RC; Shim, JP; Jones, MC	Factors influencing corporate web site adoption	INF& MAN	2001	43	article
Penn	Jansen, BJ; Spink, A; Saracevic, T	Real life, real users, and real needs	IP&M	2000	331	article
Pitt	Wade, M; Hulland, J	Review: The resource-based view and information systems researchresearch	MIS QUART	2004	109	review
Denmk	Bjorneborn, L; Ingwersen, P	Perspectives of webometrics	SCIENTOMETRICS	2001	80	article
Rutg	Jansen, BJ; Spink, A; Saracevic, T	Real life, real users, and real needs	IP&M	2000	331	article
Sheff	Thomas, O; Willett, P	Webometric analysis of departments of librarianship and information science	J INF SCI	2000	43	article
Sing	Garud, R; Kumaraswamy, A	Vicious and virtuous circles in the management of knowledge	MIS QUART	2005	28	review
Syrac	Benaroch, M; Kauffman, RJ	Justifying electronic banking network expansion using real options analysis	MIS QUART	2000	66	article
UTA	Barua, A; Konana, P; Whinston, AB; Yin, F	An empirical investigation of net-enabled business value	MIS QUART	2004	71	review
UTor	Fischer, S; Stewart, TE; Mehta, S; Wax, R; Lapinsky, SE	Handheld computing in medicine	JAMIA	2003	89	review
UW	Saha, S; Saint, S; Christakis, DA	Impact factor: a valid measure of journal quality?	J MED LIBR ASS	2003	89	article
Wuhan	Zhou, QM; Liu, XJ	Error assessment of grid-based flow routing algorithms used in hydrological models	INT J GEO INT SCI	2002	17	article

In Table 3 we show the most productive authors, the number of publications and the percentage out of the total number of publications of the university. Note that the publication types include book reviews and editorials.

Table 3. The most productive authors

University	most productive author	no. publ	% of total publ.
UCBer	BUCKLAND, M	19	18.63%
UCIrv	KRAEMER, KL	13	18.57%
UCLA	FURNER, J	17	5.59%
CMU	ALEXANDER, J	9	9.18%
DRXL	LEWIS, AM	13	4.83%
FLST	MCCLURE, CR	37	13.65%
GTEC	RENFRO, C	30	18.99%
HUMB	SEADLE, M	15	29.41%
ILUC	FAIRCHILD, CA	346	37.61%
INDB	CRONIN, B	78	17.29%
UMLND	DOPP, BJ	31	7.99%
UMBC	HOLDEN, SH	3	33.33%
Umich	SEEMAN, C	13	5.12%
UNC	KUHLMAN, JR	53	12.99%
UNT	GREISDORF, H	8	10.00%
Penn	LUMPKINS, CL	56	9.52%
Pitt	SPINK, A	25	7.06%
Denmk	HJORLAND, B	40	27.59%
Rutg	MAXYMUK, J	109	28.09%
Sheff	FORD, N	29	14.08%
Sing	PAN, G	3	42.86%
Syrac	NICHOLSON, S	21	11.05%
UTA	PETERS, SL	49	12.28%
UTor	DILEVKO, J	28	11.07%
UW	SZATMARY, D	82	16.91%
Wuhan	QIU, JP	5	11.63%

The most productive author was C. A. Fairchild, however one must note that 315 out of the 316 publications were book reviews. These book reviews were written by Constance A. Fairchild a reference librarian at the University of Illinois Library in Urbana-Champaign. Her reviews constitute 37.61% of the total publications of the University of Illinois in the category "information science and library science". Thus we see that some of the more visible authors are not from the iSchools. Next, in Table 5 we present the journals with the largest number of publications for each university.

Library Journal was by far the most productive journal in 14 cases. We see here some local effects, two of the European institutions (Sheffield and the Royal School) published extensively in the British Journal of Documentation, while faculty from the University of Toronto showed a preference for the Canadian Journal of Information and Library Science.

Table 4. Most productive journals

Univ.	most prod. journal	# publ journal	% of total publ.
UCBer	LIBRARY JOURNAL	11	10.78%
UCIrv	INFORMATION SOCIETY	13	18.57%
UCLA	LIBRARY QUARTERLY	74	24.34%
CMU	COLLEGE & RESEARCH LIBRARIES; INFORMATION SYSTEMS RESEARCH	13	13.27%
DRXL	LIBRARY JOURNAL	106	39.41%
FLST	LIBRARY QUARTERLY	52	19.19%
GTEC	LIBRARY JOURNAL	100	63.29%
HUMB	LIBRARY HI TECH	20	39.22%
ILUC	LIBRARY JOURNAL	442	48.04%
INDB	LIBRARY JOURNAL	107	23.73%
UMLND	LIBRARY JOURNAL	53	13.66%
UMBC	GOVERNMENT INFORMATION QUARTERLY	3	33.33%
Umich	LIBRARY JOURNAL	40	15.75%
UNC	LIBRARY JOURNAL	66	16.18%
UNT	PROCEEDINGS OF THE ASIST ANNUAL MEETING	26	32.50%
Penn	LIBRARY JOURNAL	157	26.70%
Pitt	LIBRARY JOURNAL	74	20.90%
Denmk	JOURNAL OF DOCUMENTATION	38	26.21%
Rutg	LIBRARY JOURNAL	147	37.89%
Sheff	JOURNAL OF DOCUMENTATION	35	16.99%
Sing	MIS QUARTERLY	2	28.57%
Syrac	LIBRARY JOURNAL	33	17.37%
UTA	LIBRARY JOURNAL	64	16.04%
UTor	CANADIAN JOURNAL OF INFORMATION AND LIBRARY SCIENCE	50	19.76%
UW	LIBRARY JOURNAL	121	24.95%
Wuhan	SCIENTOMETRICS	9	20.93%

4. CONCLUSIONS

We view this study as an exploratory study. Its limitations were clearly stated in the methods section. We recommend conducting further studies, where the searches are conducted for the individual faculty members of each of the iSchools, to get a more exact picture of their achievements. These way only publications

of the iSchools members, irrespective of the ISI category they are assigned to will be retrieved. We recommend calculating a number of measures, similar to the measures that appeared in the current study.

5. REFERENCES

[1] Adkins, D. and Budd, J. 2006. Scholarly productivity of US LIS faculty. Library and Information Science Research, 28, 374-389.

[2] Bates, M. 1998. The role of publication type in the evaluation of LIS programs. Library and Information Science Research, 20(2), 187-198.

[3] Biggs, M. and Bookstein, A. 1988. What constitutes a high quality M.L.S. program? Forty-five faculty members' views. Journal of Education for Library and Information Science, 31(1), 3-24.

[4] Boyce, B. R. and Hendren, C. 1996. Authorship as a measure of the productivity of schools of library and information science. Journal of Education for Library and Information Science, 37(3), 250-271.

[5] Brace, W. 1992. Quality assessment of library and information science school faculties. Education for Information, 10(2), 115-123.

[6] Budd, J. 2000. Scholarly productivity of US LIS faculty: An update. Library Quarterly, 70(2), 230-245.

[7] Budd, J., and Seavey, C. A. 1996. Productivity of US library and information science faculty: The Hayes study revisited. Library Quarterly, 66(1), 1-20.

[8] Cronin, B., and Meho, L. I. 2006. Using the h-index to rank influential information scientists. Journal of the American Society for Information Science and Technology, 57(9), 1275-1278.

[9] Cronin, B. and Overfelt, K. 1994. Citation based auditing of academic performance. Journal of the American Society for Information Science, 45(2), 61-72.

[10] Danton, J. P. 1983. Notes on the evaluation of library schools. Journal of Education for Librarianship, 24(2), 106-116.

[11] Egghe, L. 2006. The theory and practice of the g-index. Scientometrics, 69(1), 131-152.

[12] iConference2010 – Call for participation. 2009. http://www.ischools.org/images/bannerimage/cfp3.pdf

[13] Hayes, R. M. 1983. Citation statistics as a measure of faculty research productivity. Journal of Education for Librarianship, 23(3), 151-172.

[14] Hirsch, J. 2005. An index to quantify an individual researcher's scientific output. Proceedings of the National Academy of Sciences of the United States, 102 (46), 16569-16572.

[15] Meho, L. I., and Spurgin, K. M. 2005. Ranking the research productivity of library and information science faculty and schools: An evaluation of data sources and research methods. Journal of the American Society for Information Science and Technology, 56(12), 1314-1331.

[16] Meho, L. I., and Yang, K. 2007. Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus Scopus and Google Scholar. Journal of the American Society for Information Science and Technology, 58(13), 2105-2125.

[17] Mulvaney, J. P. 1992. The characteristics associated with perceived quality in schools of library and information science. Library Quarterly, 62(1), 1-27.

[18] Oppenheim, C. 1995. The correlation between citation counts and the 1992 research assessment exercise ratings for British library and information science university departments. Journal of Documentation, 51(1), 18-27.

[19] Oppenheim, C. 2007. Using the h-Index to rank influential British researchers in information science and librarianship.2006. Journal of the American Society for Information Science and Technology, 58(2), 297-301.

[20] Pettigrew, K. E. and Nicholls, P. T. 1994. Publication patterns of LIS faculty from 1982-1992: effects of doctoral programs. Library & Information Science Research, 16(), 139-156.

[21] Sanderson, M. 2008. Revisiting h measured on UK LIS and IR academics. Journal of the American Society for Information Science and Technology, 59 (7), 1184-1190.

[22] Seng, L. B. and Willet, P. 1995. The citedness of publications by United Kingdom library schools. Journal of Information Science, 21(1), 68-71.

[23] Shaw, D., and Vaughan, L. 2008. Publication and citation patterns among LIS faculty: Profiling a "typical professor". Library and Information Science Research, 30, 47-55.

[24] Varlejs, J. and Dairymple, P. 1986. Publication output of library and information science faculty. Journal of Education for Library and Information Science, 27, 71-89.

[25] Wallace, D. P. 1990. The most productive faculty. Library Journal, 115(8), 61-63.

[26] White, H. S. 1981. Perceptions by educators and administrators of the ranking of library school programs. College & Research Libraries, 42(3), 191-202.

[27] White, H. S. 1987. Perceptions by educators and administrators of the ranking of library school programs – An update and analysis. Library Quarterly, 57(3), 252-268.

Net Generation in Organizations: Perceptions and Strategies

Karine Barzilai-Nahon
University of Washington
karineb@uw.edu

Robert Mason
University of Washington
rmmason@uw.edu

ABSTRACT
This paper reports on an exploratory study of how executives in organizations perceive the entrance of the "*net generation*" into the workplace. We approached this question by collecting data from interviews, focus groups, and an online survey. The paper discusses the different organizational mechanisms and strategies executives use to address perceived tensions as the Net Generation enters the workforce. Particularly, we discuss executives' preference for top-down strategies and their tendency to address the triad of technology-values-behavior as separate components instead of a unified concept.

General Terms
Human Factors, Theory, Management

Keywords
Net Generation, Values, Organizations, Digital Natives, Strategies, Change

1. INTRODUCTION
Knowledge work comprises—and likely will continue to comprise—most of the value creation in the developed world. Observers note that members of the generation just now coming into the workforce as knowledge workers have grown up in a world surrounded by technologies and digital tools that enable a wider range of communication possibilities and greater connectivity than ever before in the developed world and even in developing economies. Researchers label this generation born between 1978 and 1994 as the *Net Generation* because of their perceptions of this generation as immersed in a digital environment [1], and the members of this group as "net geners." We are using this label because it is useful to highlight a portion of the generation that has been actively engaged in the digital world. However, we acknowledge that within this age group, there are individual differences in characteristics and different experiences with the range of information and communications technologies, particularly across different economies and social groups.

Some researchers and observers claim that members of this generation have developed skills, habits, and behavioral norms of using technology that differ from those of previous generations, particularly the baby boomers [1-4]. In this study we are not going to resolve the controversial claims that the *net generation* has or has not developed distinct values and behaviors. Instead, we are interested in understanding the dynamics of the entry of this generation into existing organizations. Therefore, we undertook a study that analyzes CIOs' responses to their experience and perception with the *net generation* workers.

To do this, we reviewed the extensive literature on the *net generation* (variously referred to as gen Y, net natives, digital natives, and millennials) to get an idea of the behavioral differences that might be observed by the executives as this generation entered the workforce. We synthesized these observations and research findings in a scenario and used the scenario to elicit reactions from executives. We used three different methodologies to present the scenario and collect data: interviews, focus groups, and an online survey.

2. THE CHANGING WORKFORCE
The U.S. workforce will change over the next ten years as the demographics of the population change. Demographics of the workforce are changing world-wide, but our discussion focuses on the U.S. The significance of these changes is that a large portion of the workforce (the baby boomers) will be entering retirement age soon—leaving the workforce—just as members of the *net generation* will be entering the workforce. Table 1 compares the relative numbers of the three generations to demonstrate these changes. In 2008, the Baby Boomers group comprised over 40 percent of the U.S. labor force [5]. By the year 2018, all but the youngest of this generation will be at retirement age.

Table 1. Population Estimates of Three Generations of Workers

Generation	Birth Year	Current Age in 2009	Population Estimates *
Baby Boomers	1946-1964	45-63	82.8 M
Generation X	1965-1977	32-44	50.9 M
Net Generation	1978-1994	15-31	69.1 M

* Population estimates based on US 2000 Census [6]

Note in Table 1 that the *net generation*, while not as numerous as the baby boomers, has about 36% more members than generation X. Prensky estimates members of this generation will have spent over 10,000 hours playing videogames, sent and received over 200,000 emails and instant messages, spent over 10,000 hours talking on cell phones, and over 20,000 hours watching television *before* they even graduate from college (e.g., before they reach 21 or 22 years of age, about when they might be entering the workforce) [2]. A large majority of teens in the United States (over 90%) use the Internet [7] and over 71% of teens use mobile phones [8]; both play a major role in their relationships with their friends, families, and schools.

3. COMPARING TWO GENERATIONS
Because of the size of the *net generation*, considerable research already exists on how its members play, learn, and work. Marketers, educators, corporations, and employers recognize the need to understand the *net generation*'s learning and working styles. The Pew surveys have examined the changing uses of communications technology and the accompanying changes in values with younger generations [9-11]. Others have used these and other studies to reach different perceptions about how the *net generation* thinks and behaves. In the case of Twenge [4] and Tapscott [1], they go further. They claim that the *net generation* is not only perceived as different from the baby boomers but they

are actually distinctly different from them in values and behavior. However, they do not agree on the significances of these differences.

Abram and Luther [12] identified nine aspects of the *net generation* behavior that they believe differentiate this group from its predecessors. Additionally, they claim that members of the *net generation* exhibit fundamental differences in the use of information, personal interactions, and social values. Among the distinguishing aspects are multitasking, experiential, collaborative, adaptive, and direct behaviors.

Table 2 compares the set of values, attitudes, and styles of the *net generation* and baby boomers as perceived in the literature [1, 4, 9-12]. Many of the differences highlighted in this table can serve as the genesis for potential issues and tensions as members of the *net generation* join organizations.

Table 2: Perceived Differences in Behaviors and Values

Behaviors and Values	***Net Generation***	***Baby Boomers***
Work Style	Multitasking	Time management
Learning Style	Learn from experience	Learn from instruction
Collaboration	Collaborative	Independent
Motivations	Positive reinforcement	Competition
View on Authority	Respect for others is earned	Respect for authority
Structure	Decentralized, non-hierarchical, inclusive	Centralized, hierarchical, exclusive
Information Access	Access for all	Access to those in power

4. METHODOLOGY

Our methodology in this study comprised three steps: scenario development, data collection (interviews, focus groups, and an online survey), and thematic analysis of the data.

First, based on the literature review, we developed a scenario that reflected some of the potential issues between executives and younger generations as shown in table 2. The scenario was developed to address a target audience of Chief Information Officers (CIOs), Chief Technical Officers, and other executives in companies that use information technology extensively and who might be hiring members of the *net generation* for their organizations (see the scenario in appendix).

Second, we collected data through three methods: interviews, focus groups, and an online survey. For our interviews, we used a snowball technique beginning with executives who served as advisors to a Master of Science in Information Management program. We identified ten CIOs and CTOs from government and for-profit organizations, sent them the scenario, and conducted interviews that lasted 20-40 minutes. For our focus groups, we used convenience sample of 110 CIOs, CTOs, and other executives that were attending a seminar on managing the information technology function. We had 12 groups of 8-10 persons, and each group had a moderator and note taker. The participants represented a variety of businesses, from engineering firms to health care. For the online survey, we posted the scenario to a website, announced the study to executives who were subscribers to a consulting service, and received 49 responses.

In each of these three methods we presented the scenario and asked the participants to respond to four questions:

1. Are there any issues that you've experienced or observed that are missing from the scenario?
2. What issues do you feel are most critical at this point in time?
3. How are you addressing the issues identified in question 2?
4. Do you see some issues as becoming more important over time?

The research team (the authors and two research assistants) reviewed and coded the transcribed interviews and focus group notes using thematic analysis. We identified concepts to identify specific organizational responses to the issues and grouped the responses into clusters of mechanisms used by executives and organizations.

5. RESPONDING TO THE NET GENERATION: ORGANIZATIONAL MECHANISMS

In this paper we are reporting on the answers to question no. 3 only (see the above section on the methodology). The data for the thematic analysis we conducted for understanding how executives address perceived tensions were collected from responses to the question: 'How are you addressing the issues you identified as most critical?' We categorized responses into four main clusters of organizational mechanisms: project management, technology, human resources and policy.

Project Management

The project management responses focused on defining management rules, testing performance, and restricting ways of working. One example of the project management approach is the following:

> *"…whether the employees of the company want to or not, in order to be effective as a full team they've got to work in a similar manner. The organization put together for all of our core activities a series of execution procedures that we follow in order to make sure that we are as productive as we can be. Some of the kids when they come in don't necessarily want to change, you know they think in some ways it's an old way of doing things…"* [I-3]

Technology

The technology mechanism responses referred to the use of technology to address tensions, often taking the form of prohibiting or restricting the use of particular technologies. Examples include:

"We've restricted instant messaging and blogs. And until we get another fight years down the road we're not going to open up instant messaging. …when we do open up instant messaging it will be for internal communications only. …What we are trying to do is provide business tools to perform business functions for business solutions, so when people come in, you know we make them sign all the usual security agreements and tell them that the technology tools in the company are for business purposes…occasional personal use is fine…but instant messaging and anything with blogging or chat rooms or anything like that isn't acceptable to the company." [I-3]

"…we don't allow IM in our equipment firewall…" [I-4]

Human Resources

The human resources mechanism responses referred to the use of the HR function in addressing the tensions, either early (to improve screening and hiring for fit) or later in policies and training. Examples of these responses include:

"So, I didn't get the attorneys involved but… I did need HR's perspective on the trade-off of taking away from the employees something they knew they could be doing and what might it demonstrate in terms of the corporate attitude…" [I-2]

"…we have been trying to do this mostly by training people we have had through HR and through our legal department we are trying to have information meetings…" [I-7]

Policy

In this cluster, we collected responses that referred to organizational processes, managing risk, and specific organizational policies. Examples include:

"…we have a policy so it starts with a policy around, 'you know, our business tools are meant for business reasons'..." [I-5]

"… we produced a policy statement and sent it out to all of the employees, which unfortunately was written in policy-eze language vs. a more warm and friendly memo, and it pointed out that all of the corporate assets, including our computers and phones, etc, etc were for the use of the employees at work, some reasonable amount of personal use was allowed, but, and then we itemized the types of things they weren't supposed to be doing." [I-2]

When examining carefully the four organizational mechanisms that emerged from the thematic analysis, we observe that each of them also can be mapped in terms of types of management strategy applied by the executives [13]. These strategies differ in terms of the *net generation*'s and executives' involvement, the decision approach of the executives, the duration and scope of the change, and the implications for resources [14]. Table 3 shows this range of strategies and the percentage of executives using each. Note that because executives use multiple and mixed strategies according to different situation, therefore the sum of their responses totals to more than 100%. We further found that some executives prefer instead of adopting one of the strategies in Table 3 to "wait and see" and not take any actions until it is necessary.

Table 3 - Types of Strategies and Frequency of Use

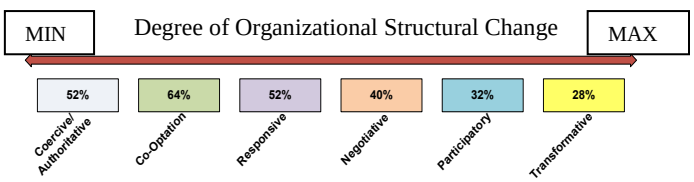
Strategy	**Description**	**% using this strategy***
Coercive/ Authoritative	"It is my way or the highway." In this strategy the organization prefers to enforce existing policies with minimal changes. This strategy is one-sided and top-down driven.	52%
Cooptation	"Manipulative." In this strategy the organization influence and manipulate employees from the *net generation* to accept the existing organizational culture and policies through different mechanisms (e.g., socialization). This is less direct, but still a one-sided and top-down driven strategy. It may involve ostensible participation, but the goals and results are similar to the coercive strategy.	
Responsive	"Flexible firefighting." This is a deliberate strategy that reacts to individual issues as they arise. The choices are context sensitive; the decisions are based on tradeoffs made unilaterally by the executives' assessment of the costs and benefits of different alternatives.	52%
Negotiative	"Making compromises." In this strategy executives negotiate and make tradeoffs on critical issues with the participation of the *net generation*.	40%
Participatory	"Let's play together." This strategy involves full engagement and collaboration by all stakeholders in the organization's vision and operational processes.	32%
Transformative	"Melting Pot." In this strategy the organization changes its structure and norms to something new.	28%

*% refers to percent of executives' (N=160) responses in the named strategy classification. Since respondents can use multiple strategies, the total is >100%

6. DISCUSSION

The results we presented above illuminate the growing awareness of executives on the recurring nature of the tensions with members of the *net generation* and with use of the newer technologies in general. This growing awareness causes them to address the tensions in a more systematic way. This is particularly evident when looking at the strategy preferences of executives, who choose strategies that are not dependent on the particular situation (as in the 'responsive' or 'negotiative' strategies) as shown in Table 3. These strategies can be mapped along an axis corresponding to the degree of organizational structural change (changes in decision-making and power relationships) required for implementation (see Figure 2).

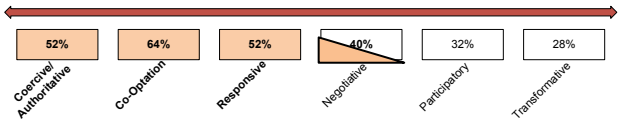
Figure 1: Degree of Organizational Structural Change



We would like to discuss two phenomena we observe in our findings: 1) the priority executives give to top-down strategies as opposed to bottom-up ones. 2) the preference of executives to control either behavior or technology determinants while ignoring values and norms, which we believe form the third apex of an integral triad.

6.1 Choosing Top-Down Strategies as a Priority

Figure 2: Top-Down Strategies



In Top-Down strategies, executives dictate the boundaries, goals, and, to a large extent, the outcomes. Figure 3 illustrates the prevalence of top-down strategies for dealing with the *net generation*: the Coercive, Cooptation, Responsive and, to some degree, the Negotiative. Here are two quotes that exemplify the top-down approach:

> *"Must set very clear goals/expectations. Need to manage and micro-manage more than with previous generation of employees. Need more mentoring by senior people to train new employees on how to produce high-quality outputs."* [S-9]

> *"Training is key, and setting expectations correctly at the time of hire."* [S-17]

Management literature suggests that top-down strategies may be ineffective in dealing with changes in an organizational context specially for the long-term [15]. This could apply to the *net generation* as well, which may require organizations to perform some changes on their behalf. In the long-term, top-down strategies have the potential to stimulate higher levels of resistance to attempts at control, especially in periods of change [16, 17]. Conversely, creating and maintaining a cohesive organizational culture in a process that involves all stakeholders has higher chances for long-term success [18, 19]. In the near term, a top-down strategy can alienate the younger employees, decreasing the chances to build a shared and common vision, mission, and organizational culture and increasing turnover.

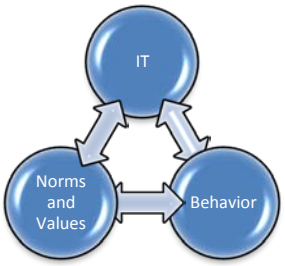
> *"So I think it's [change] got to be on both sides. If you don't, you won't have staff. I mean, I don't think corporations who stick to the old way of doing things are going to be able to maintain any kind of staff base unless they adapt or are willing to hire people with other work ethics."* [I-4]

Finally, addressing challenges in a top-down manner often requires dictating behavior uncommon to the *net generation* members. This is an example of treating the symptoms and not the underlying cause. The *net generation* initially might be compliant, but the gaps in behaviors and values remain; Organizational behavior literature agrees that gaps in behavior and values in most cases create a dissonance, that later is translated into the need for change [20]. Leaders are likely to find they need to address the same fundamental challenges again unless they are resolved at a more fundamental level.

6.2 The TVB (Technology-Values-Behavior) Triad

A "generation gap" is not a new phenomenon. The values and behavioral norms of succeeding generations have always differed in some degree from past ones. Also, it is generally accepted that information technology shapes many organizational norms, values and behavior, and the reverse is also true [21]. Additionally, groups take technology and appropriate it to their own needs. None of this is new. What is new is the extent, timing, speed and the closeness of this recursive relationship between information technology and the *net generation's* values and behavior. We believe that understanding and resolving the tensions arising from perceptions about the *net generation* can only be achieved if we use a lens that considers technology, values, and behavior as a closely coupled triad of factors affecting the perceived organizational tensions.

Figure 3: The Technology-Values-Behavior Triad



One of the things we observed in the data is that executives in many cases seek to control either behavior or technology determinants to resolve tensions. Decomposing this triad into separate components and trying to resolve issues by treating only one component at a time may not be effective due to the close relationship between these concepts. We posit that this triad should be treated from a holistic point of view. One of the consequences of the information society is that these three components move together and are closely coupled.

Executives' decomposition of the triangulation of technology, behavior and norms also helps to explain the failure of top-down strategies, which inherently focus on regulating behavior either through rules and policy or technology. It is not a coincidence that most CIOs chose top-down strategies to address tensions resulting from their entry into the workplace. These strategies require minimum critical structural and political changes to the organization because the compromises to operational processes are typically minimal.

We also observe that executives approach the behavior of members of the *net generation* (and other behavior associated with use of the newer communications technologies) from the individual level and ignore the norms that emerge from social groups. For example, managers believe that they can train individuals to behave according to the company rules and this will solve the tensions they perceive.

We suggest that the new unit of analysis should be *communities* rather than individuals. The technology component in Figure 3 provides platforms for communities to be established quickly; these communities establish and reify norms and reinforce behaviors at a pace that has not been observed as prior generations

entered the workforce. By choosing strategies that focus only on the individual level, ignoring the complexity of the communal values interwoven with the technology use and behavior, executives will find it difficult to enforce desired behavior for the long run.

This study was designed and implemented as an exploratory study, and the methods and resulting data have the limitations associated with such studies. The sample from which data were collected was not necessarily representative of the entire universe of executives and CIOs.

Respondents in our study were predominantly from established organizations, and these are more likely to experience the tensions than a newer organization such as Google or Facebook. (Facebook was established by *Net* Geners.) Such organizations can have a different workforce demographic and may not have the legacy systems that can stimulate the tensions perceived by our respondents.

7. CONCLUSIONS

Members of the *net generation* are perceived by executive and others as using information technologies in ways that differ significantly from those of prior generations. They are also perceived as having values and behavioral characteristics that differ from prior generations. In many cases these behaviors are viewed as inefficient, ineffective, or even unethical by those already in the work force. These perceptions, whether true or not, stimulate tensions between new employees from the *net generation* just entering the workforce with other generations. Similar tensions can arise when others adopt new technologies and behave like the *net geners*.

According to the executives we interviewed, few organizations currently are set up to accommodate these behaviors. Organizations have an inertia that inhibits rapid change, and this presents a challenge even to executives who recognize the need to change. Moreover, organizations that have been led by baby boomers have processes and information systems that were designed by baby boomers, for baby boomers, using technologies that were available at the time baby boomers were becoming managers. These legacy systems, and the accompanying comfort with their use by baby boomers, add to the inertia.

However, most CIOs and CTOs recognize the challenge they will be facing as their workforce becomes more populated with members of the *net generation*, and some executives already are working to deal with the issues. For those that do recognize the issues, they are using (or planning to use) different strategies, which we discussed in this paper. It appears that most executives feel more comfortable using top-down approaches, which may not be effective to address tensions with the *net generation*. We suggest using the TVB (Technology-Values-Behavior) triad as an effective holistic lens through which researchers and practitioners should analyze the ecological system of the *net generation*. A consequence of taking this ecological view is that the concept of communities becomes embedded in strategic management practices.

8. REFERENCES

1. Tapscott, D., *Grown up Digital*. 2009, New York: McGraw-Hill Companies.
2. Prensky, M., *Digital Natives, Digital Immigrants: Do they really think differently?*

. On the Horizon, 2001. **9**(6).

3. Rainie, L., *Digital Natives: How today's youth are different from their "digital immigrant" elders and what that means for libraries*, in *PEW Internet & American Life Project*. 2006.
4. Twenge, J.M., *Generation me : why today's young Americans are more confident, assertive, entitled--and more miserable than ever before*. 2006, New York: Free Press.
5. Poulos, S. and D. Smith, *Aging Baby Boomers: In a New Workforce Development System*. 2008, Urban Institute: Washington D.C.
6. Bureau, U.S.C., *QT-P1. Age Groups and Sex: 2000. Census 2000 Summary File 1, Matrices P13 and PCT12. Retrieved 8/09/07 from: http://factfinder.census.gov/servlet/QTTable?_bm=y&-qr_name=DEC_2000_SF1_U_QTP1&-geo_id=01000US&-ds_name=DEC_2000_SF1_U&-_lang=en&-format=&-CONTEXT=qt*

2000.

7. Jones, S. and S. Fox, *Generations Online in 2009*. 2009, Pew Internet & American Life Project: Washington D.C.
8. Lenhart, A., *Teens and Mobile Phones Over the Past Five Years*. 2009, Pew Internet & American Life Project.
9. Lenhart, A. and M. Madden, *Teens, Privacy & Online Social Networks How teens manage their online identities and personal information in the age of MySpace*. 2007, PEW Internet & American Life Projects: "Washington, D.C.".
10. Lenhart, A., L. Rainie, and O. Lewis, *Teenage life online: The rise of the instant-message generation and the Internet's impact on friendships and family relationships*. 2001, PEW Internet & American Life Projects: "Washington, D.C.".
11. Lenhart, A., M. Madden, and P. Hitlin, *Teens and Technology: Youth are leading the transition to a fully wired and mobile nation*. 2005, PEW Internet & American Life Projects: "Washington, D.C.".
12. Abram, S. and J. Luther, *Born with the Chip:The next Generation Will Profoundly Impact Both Library Service and the Culture within the Profession*. Library Journal, 2004. **129**(8): p. 34.
13. Birkinshaw, J., and Jules Goddard, *What is Your Management Model?* Sloan Management Review, 2009. **50** (2).
14. Barzilai-Nahon, K. and R. Mason, *How Organizations Respond and should respond to the Next Generation of Workers--Digital Natives and Organizations (working title)*. 2009, The Information School of the University of Washington Seattle, WA. p. 18
15. Pascale, R.T. and J. Sternin, *Your Company's Secret Change Agents.* Harvard Business Review, 2005. **83**(5): p. 72-81.
16. Weber, M., ed. *Economy and Society: An OUtline of Interpretive Sociology*. ed. G. Roth and C. Wittich. 1978, University of Californai Press: Berkeley, CA.
17. Kotter, J.P. and L.A. Schlesinger, *Choosing Strategies for Change.* Harvard Business Review, 2008. **86**(7/8): p. 130-139.
18. Hill, L., *Becoming the Boss.* Harvard Business Review, 2007. **85**(1): p. 48-56.
19. Christensen, C.M., M. Marx, and H.H. Stevenson, *The Tools of Cooperation and Change.* Harvard Business Review, 2006. **84**(10): p. 73-80.
20. Schein, E., *Models and Tools for Stability and Change in Human Systems.* Reflections, 2002. **4**(2): p. 34-46.

21. Orlikowski, W.J. and D. Robey, *Information Technology and the Structuring of Organizations.* Information Systems Research, 1992. **2**(2): p. 143-169.

An Operational Definition of the Information Disciplines

Marcia J. Bates
Department of Information Studies
University of California at Los Angeles (UCLA)
Los Angeles, CA USA 90095-1520
1-310-206-9353
mjbates@ucla.edu

ABSTRACT

The author and Mary Niles Maack set out to develop the content for an up-to-date edition of an encyclopedia intended to cover all the major information disciplines in an integrated fashion. This effort arose from the belief that all the i-disciplines have core interests in common, and that the commonality needs to be brought out in the structure, organization, and content of the encyclopedia—without denying the inherent differences between the fields as well. This plan led to four years of effort in designing the encyclopedia, recruiting authors for its entries, and organizing the resulting array of entries in a classification that reflects the major topical areas of the information disciplines. The resulting design of the encyclopedia can be seen as an operational definition of the i-disciplines. The design effort and its results are described.

Keywords

Information disciplines; i-disciplines; Encyclopedias, Operationalization of concepts; Disciplinary definitions; Social studies of information

1. A NEW ENCYCLOPEDIA FOR THE INFORMATION DISCIPLINES

The author and Mary Niles Maack contracted in 2005 to edit the Third Edition of the *Encyclopedia of Library and Information Science,* as Editor-in-Chief and Associate Editor, respectively. We did not, however, want an encyclopedia solely of traditional LIS. We felt that the information disciplines covered a much broader range than that single field, and that it was time to produce an integrated reference tool for a much wider array of information disciplines. We felt that the iSchool movement well reflected that broader range, and we set out to design the encyclopedia to be a unified expression of the full array of the information disciplines. In the process of that four-year effort, we gained a deeper understanding of the information disciplines, and found a way to integrate those disciplines into a single seven-volume encyclopedia, which appeared at the end of 2009 [1]. The resulting design of the encyclopedia can be seen as an operational definition of the i-disciplines. (Conceptual definitions describe a phenomenon in principle; an operational definition describes the particulars that will stand for that conceptual definition in a specific situation.) This article describes that effort and its results.

In fact, we would have liked to name it the *Encyclopedia of the Information Disciplines,* and may do so in the future. We believe, however, that this edition is transitional, that buyers and readers accustomed to the prior *ELIS* title need to see how LIS can be integrated with the other information disciplines. At the time we began, "information disciplines" was not a widely-used phrase, and we did not want to lose readers through the use of a title so unfamiliar. We did, however, make the title plural: the *Encyclopedia of Library and Information Sciences,* to better reflect the range of coverage.

In our invitations to prospective writers for the encyclopedia, we described the effort as follows:

> We are endeavoring to make the forthcoming Third Edition into an authoritative guide to the 21st century information disciplines—we're including informatics, information systems, knowledge management, archives, records management, museum studies, bibliography, document and genre studies, and social studies of information, along with LIS. We are working with the assistance of a 50-person international Editorial Advisory Board of premier researchers and practitioners from all these domains. We believe this online and multi-volume print edition will constitute a substantial addition to the literature of all the information sciences.

The encyclopedia has just been published.[1] It contains 565 article-length entries, ranging from 1000 to over 20,000 words each, averaging between 5,000 and 8,000 words. My objective here is to describe not so much the practical sequence of developing the contents list, but rather the intellectual process we went through. (The author was responsible for most of that effort, with invaluable input from the Associate Editor.)

2. MAJOR ISSUES IN THE DEVELOPMENTAL PROCESS

Three principal issues arose in the development of the contents list for the encyclopedia, discussed below. A fourth issue, raised by a reviewer, will be addressed below as well.

2.1. Professional or Academic?

The list of fields given above has a strong professional tilt; as a practical matter, many information disciplines are most visible in society in their professional manifestations. But we in academia, especially, know well that the information disciplines also contain a rich body of research and theory that is growing larger by the day. The information disciplines are most assuredly also *academic*. Furthermore, no profession is truly a profession without a body of theory underlying and infusing practice with a deeper meaning than quotidian transactions would imply on their own. In a word, the answer to the above question is: Both. The encyclopedia addresses both research and practice in ways to be described below. There is surely no longer any question that there is a body of theory in the information disciplines both driving professional practice and providing enrichment to many other academic disciplines in a purely intellectual exchange.

Information *professions* concern themselves with gathering, evaluating, organizing, storing, retrieving, and making available information to users. Information as an *academic body of research and theory* is concerned with 1) describing and understanding the universe of recorded information of all kinds (the physical question), 2) studying human beings seeking and interacting with information in all contexts (the social question), and 3) putting people together with information by means of information technology (the design question) [2].

Bottom line: the information disciplines are both academic and professional; both theory and practice, and those dimensions are reflected in the design of the encyclopedia.

2.2. Sciences or Humanities?

The author has argued elsewhere [3] that the information disciplines are meta-disciplines; they cut across the entire conventional academic spectrum that ranges from the arts and humanities, through the social sciences, natural sciences, and mathematics and logic. The content of the information stored and organized by practitioners of the information disciplines can range across all kinds of recorded knowledge, all subject matters. Just as educators teach all kinds of subject matter, and journalists pursue news in every domain of life, so also do people in the information disciplines address all kinds of information. *The meta-disciplines shape the subject matter of all the traditional disciplines according to the social purpose of the meta-discipline*. In that sense we range across all the conventional disciplines. See Figure 1 below.

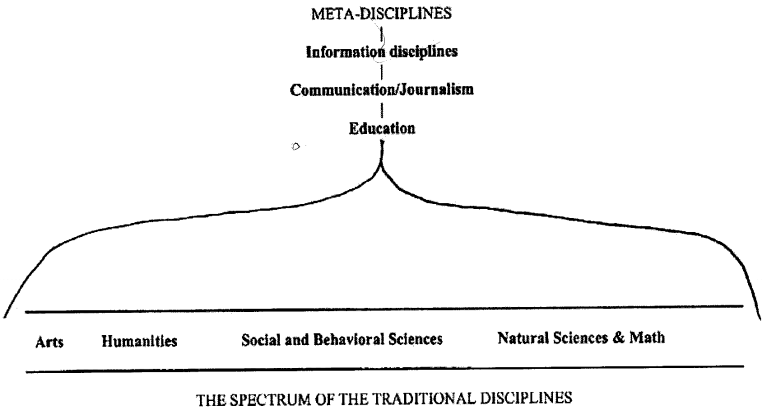


Figure 1: The meta-disciplines shape the subject matter of all the traditional disciplines according to the social purpose of the meta-discipline.

What distinguishes the information disciplines from the conventional academic spectrum, is that *whatever the information content,* the information disciplines' objectives are to understand the domain of information and the human relationship to it at a theoretical level, and to develop practical skills of organization, retrieval and dissemination of information to address real-world problems at a professional level.

For historical reasons, almost all of the disciplines on the conventional academic spectrum have some connection with the information disciplines. Often, the first information work in a field is done by people in that field, in response to pressing needs for organizing and retrieving information in that discipline. As was noted elsewhere:

> From where, then, do the new information disciplines arise? The fundamental engine of development is *need*. Human beings want to retain informational resources, and, after a very short time, these resources collect at such a rate that some principles of selection, organization, etc., need to be brought to bear, in order for the resources to continue to be available for effective use. (In earlier centuries, the rate of collection was so slow that the need was not as easily seen or acted upon. That situation has changed dramatically in the twentieth and twenty-first centuries.) As resources collect, interested individuals recognize the problems and then resolve them through theoretical and professional development of ideas and practices. Those individuals either draw upon earlier information disciplines or invent or re-invent solutions to their problems.
>
> In almost all these cases, however, the interested individuals come out of one or more of the traditional academic or professional disciplines. Drawing on these individuals' education and experience, a perspective, a cognitive style, an emotional tone, and a body of knowledge are all brought to bear on the informational problem. Thus, for example, the need to organize historical archives was first tackled, usually, by historians. The need to store and retrieve radiological records first became known to medical personnel, and thus members of the medical professions were among those who first attacked the problem of radiology informatics. As a consequence, the writing and thinking in archival theory is strongly humanities-oriented in character, while a more technical approach, seated, above all, in the needs of medicine, drives radiology informatics. [3]

If we look across the full range of disciplines covered in the encyclopedia, it soon becomes evident that both scientific and humanistic forms of research and practice are to be found in the several fields. Participants in the several information disciplines have brought their training and cognitive styles to bear on information questions. For example, archives, museum studies, and bibliography tend to draw on strong humanities modes of thought, while science and engineering clearly drive informatics, information systems, etc. Thus, to label properly the i-fields addressed in the encyclopedia requires us to refer both to *the information sciences* (roots in the sciences) and to *the disciplines of the cultural record* (roots in the humanities). (Other information disciplines display a social science approach to their research areas, as in LIS, records management, and social studies of information. The social sciences, in turn, draw upon both scientific and humanities cognitive styles.)

However, though scholars and practitioners may bring a certain cognitive style with them from their undergraduate or other graduate educational experiences, in the end, most develop the conceptual and research approaches that are most effective for their area of specialization in the i-disciplines. So, for example, the art historian working on improving access to arts information may discover that he/she also has a knack for systems analysis when developing an interface for arts information resources. Most people coming into the information disciplines from other fields find that their awareness and skills must broaden out to encompass a wider range of techniques and approaches than were typically used in the home discipline.

In the selection of topics and of prospective authors for the encyclopedia, we included people with all of the various academic backgrounds and cognitive styles reflecting both scientific (nomothetic) and humanities (idiographic) styles of research, thinking, and writing.

2.3. Why Museum Studies?

It is likely that the most controversial decision in the development of the encyclopedia was the choice to include museum studies. Elsewhere [3], the author argues this in greater detail. Suffice it to say here that museums of all kinds select and display works of art and historical and archaeological artifacts of our cultural heritage. They display the results of our biological and geological research in science museums. Thus, museums, too, range across all forms of knowledge, and are concerned with evaluating, collecting, storing, and making available both heritage materials and scientific knowledge. Furthermore, with the growth in collections of online images, many museums are developing a digital presence as great as that of any library or archive. These commonalities have been recognized by the U.S. federal government in the creation of the granting agency, the Institute for Museum and Library Services (IMLS).

2.4. Relation to Computer Science?

One of the reviewers argued that the encyclopedia was traditional because it diminished the role of computer science. I could not disagree more, but am glad that this issue was raised, as it had not been addressed explicitly in the earlier draft. For all sorts of good and necessary reasons, computer science looms very large in most of our iSchools. Nonetheless, it is *not* the case that "iSchool = computer science school." From the beginning, it has been understood that iSchools represent a new departure in a world of growing information technology and IT use; iSchools are not simply sites of computer science applications, or branches of computer science departments.

What then is it that iSchools teach? Because computer science is so important in the world and in the universities today, it is often the only part of the information disciplines that is seen and readily understood by people outside our field. Furthermore, computers are concrete and visible in most people's lives, while information is less tangible and less well understood as a phenomenon of direct study and work-related activity. There is a real risk, consequently, that computer science and its various branches will come to be seen as virtually identical to the i-disciplines, and will drive the development of the iSchools in universities.

But there is a reason that iSchools have the name they have. There is a unique perspective that the information disciplines bring to the table. Computers are vital to all modern information disciplines, but so is an understanding of information technology *from the standpoint of the information and the information users,* not just from the technology side. iSchools are called what they are because the people that graduate from them understand how information operates in society and in people's minds, they understand how to design information systems to respect the nature/structure of the information, and the nature of the human patterns of use of them, and the larger social and cultural context that embeds that information technology. Good information access, retrieval, and use must also be designed from these other perspectives to be optimally effective in society. Computer science may be the engine, but it is not the whole automobile, nor the automobile drivers and riders, nor the system of roads—all of which must be understood and designed and integrated into a single effective system.

There are many encyclopedias of computer science. It was an encyclopedia of the information disciplines that we needed, and *ELIS* is intended to fulfill that need. *ELIS* does, in fact, deal with computer science in many ways, but it does so by looking at computer science from the perspective of the information disciplines, rather than looking at the information disciplines from the perspective of computer science.

3. THE FINAL STRUCTURE OF THE ENCYCLOPEDIA

3.1. Rejecting "Silos"

The Introduction to the *Encyclopedia* notes the following:

> Both the Editors and the Editorial Advisory Board agreed that we did not want parallel "silo" groupings of entries on the several disciplines. The purpose of this edition is to address these related disciplines in a way that both demonstrates the unities across the fields, as well as recognizes their uniquely distinguishing characteristics. Thus, the choice of topics reflects this persistent duality; some authors address a topic across the disciplines, other authors specialize in what they know best. In many cases, but not all, what has been learned in one field can be applied in others. Fund-raising techniques used in one non-profit area can usually be utilized in another non-profit area. On the other hand, only librarians are likely to need information about serials vendors, and only museum professionals have to address trafficking in art objects.[1]

Making the decision not to create parallel sets of entries identical for each component discipline was an interesting challenge, and resulted, I believe, in much richer and more integrated contents for the encyclopedia than would have been brought about otherwise. Editorial Advisory Board members were invaluable to this process, providing many topics from their areas of expertise. For this Editor, developing the contents list felt like a snake shedding skins. Dozens of distinct spreadsheets were developed in the process, and each type of spreadsheet went through multiple versions.

3.2. A Different Way of Categorizing

Disciplines, especially ones with a professional element, are much more complex than may at first appear. Developing the encyclopedia contents was not just a matter of writing down various possible topics and then organizing them. Disciplines are also composed of people, organizations, standards, histories, projects, etc., and decisions must be made about all of these. Both because of the availability of some other biographical resources, and a dearth of historians writing on the many individuals that we might want to profile, we chose early on to exclude biographies of people. However, that still left many other components of these several disciplines.

In the end, we identified seven broad classes of material to include, subdivided into major sub-categories.

• Descriptions of the information disciplines and sub-disciplines themselves (Category 1).

• The core *research and theory* of the information disciplines (Categories 2 and 3).

• Those internal conceptual components that are distinctive to the information disciplines: 1) the information technology, 2) the information itself, and 3) the human beings using information (Categories 5, 6, and 8).

• The specifically *professional* skills and activities that information professionals engage in (Category 7).

• The social and economic infrastructure supporting the disciplines and professions: Institutions; and Organizations (Categories 4 and 9).

• The geographical and political dimensions: The national cultural institutions and resources of countries (Category 10).

• History (Category 11).

Figure 2 displays the broad conceptual structure of the final encyclopedia, with categories numbered 1 to 11. Both the online and paper forms of the encyclopedia will of course be alphabetical, but they will also contain a "Topical Contents List," which groups all related entries together under the final chosen rubrics of the list in the figure. We do not have the space here to reproduce the entire contents, but an example entry title or two is given for each category in the topical list in Figure 2.

1. Information Disciplines and Professions
--General Disciplines *Information Science; Knowledge Management*
--Disciplinary Specialties *Biomedical Informatics; Music Librarianship*
--Cognate Disciplines *Artificial Intelligence; Epistemology*
--Career Options and Education *Careers and Education in Information Systems*
2. Concepts, Theories, Ideas
--Key Concepts *Information; Literacy; Work of Art*
--Theories, Models, and Ideas *Classification Theory; Information Society; Sense-Making*
3. Research Areas
--Cross-Disciplinary Specialties *Information Arts; Linguistics and the Information Sciences*
--Research Specialties *Webometrics; Personal Information Management*
4. Institutions
--Generic Institutions *Archives; Libraries; Museums*
--Institution Types *Corporate Archives; School Libraries*
--Named Institutions *Library of Congress; Louvre*
--Ancillary Cultural Institutions *Art Galleries; Historical Societies*
--Collections *Rare Book Collections; Test Collections*
5. Systems and Networks
--Information Systems *Knowledge Management Systems; Search Engines*
--Network and Technology Elements *Data Transmission Protocols; Intranets*
6. Literatures, Genres, and Documents
--Literatures *Grey Literature; Economics Literature History*
--Generic Resources *Digital Images; Folksonomies; Markup Languages*
--Named Resources *OAIS Reference Model; Resource Description Framework (RDF)*
7. Professional Services and Activities
--Appraisal and Acquisition of Resources *Digital Content Licensing; Museum Collecting*
--Institutional Management and Finance *Archival Management and Administration*
--Organization and Description of Resources *Cataloging; Museum Registration*
--Resource Management *Oral History; Serials Management; Version Control*
--User Services *Reference Services; Exhibition Design; Storytelling*
8. People Using Cultural Resources
--General *Reading Interests; Visitor Studies*
--Population Groups *LGBT Information Needs; Older Adults' Information Behavior*
--Subject Areas *Social Science Literatures and Users; Engineering Literatures and Users*
9. Organizations *Bibliographical Society; International Council on Knowledge Management*
10. National Cultural Institutions and Resources *Japan: Libraries, Archives, and Museums*
11. History *History of Records and Information Management; SMART System, 1961-1976.*

Figure 2: Main Topical Categories of *ELIS*, with some example entry titles.
Italicized topics are examples in the stated category. Total encyclopedia entries: 565.

4. DISCUSSION

In line with the issues raised above regarding the selection of topics for the encyclopedia, note the following further points about the categories:

4.1. Disciplines and Sub-fields

The topics in the first category cover entries about the whole disciplines, their sub-fields, and education and careers in those fields. One of our hopes is that readers will discover how many disciplines and sub-disciplines there are in the information fields; indeed, they may be made aware of some other disciplines for the first time. We also include entries on a penumbra of cognate disciplines that have strong information components, but which may not see themselves as information disciplines per se. [Total: 76 entries.]

4.2. Research and Theory

Key concepts, models, and theories are described in categories 2 and 3, as are social and cultural issues associated with the several disciplines. In particular, iSchools have demonstrated strong interest in the social and policy aspects of information. See also discussion of Figure 3 in next section. [Total: 130 entries.]

4.3. Areas Distinctive to the Information Disciplines

Categories 5, 6, and 8 address the information technology, the information itself, and people using information; these are the commonly mentioned heart of the i-disciplines and the iSchool curricula. It is often said that the iCaucus addresses the nexus of people, technology, and information. These three categories represent that core. In a very real sense, almost all entries in the encyclopedia deal with these three elements coming together in one way or another, each entry emphasizing one aspect or another. We believe that the nearly 30 entries on human information behavior (HIB) represent the first time an encyclopedia has addressed HIB in such detail. [Total: 111 entries]

4.4. Professional Elements

In category 7, the range of professional services and activities is addressed. It is here that the similarities and differences of the several fields are most in evidence. All the information professions share tasks that have much in common: Evaluation and acquisition of resources, management of the institution holding the resources, organization of resources, monitoring and management of resources, and user services. Librarians study "user behavior" and museum professionals conduct "visitor studies." These practices sound different, but they have much in common, and could no doubt learn from each other. As well, each profession has some areas or issues that are distinctively its own. [Total: 52 entries.]

4.5. Social and Economic Infra-structure

The Institutions and Organizations categories (categories 4 and 9) represent the physical and social manifestation of human aggregation around social tasks to be achieved and problem areas to be addressed, as these play out in the information disciplines. [Total: 114 entries.]

4.6. Geographical and Political Manifestations

A prime interest of Associate Editor Maack was to develop a set of entries that surveyed the cultural infrastructure, particularly with respect to libraries, archives, and museums, of as many countries worldwide as possible. Section 10's entries, some on countries not often described in the literature, represent a strong start. [Total: 45 entries.]

4.7. History

Finally, we wanted the history of the information disciplines to be covered, and some entries in the last section, section 11, provide the needed historical perspective on the rest of the entries in the encyclopedia. The section includes entries on some topics not often dealt with historically, such as the history of word processing. [Total: 37 entries.]

5. SOCIAL AND CULTURAL STUDIES IN THE INFORMATION DISCIPLINES

Finally, the topical component likely to be of particular interest to iConference attendees is the section on research specialties, in particular, the social, cultural and policy aspects of the information disciplines. So, Figure 3 lists the all the sub-areas under "Research Specialties," with example entry titles from those sections.

1.Bibliometrics, Scientometrics
Examples: Citation Analysis; Information Scattering

2.Information Behavior and Searching
Examples: Information Searching and Search Models; Knowledge Sharing Mechanisms

3.Information Organization and Description
Examples: Metadata and Digital Information; Moving Image Indexing; Topic maps

4.Information Retrieval
Examples: Recommender Systems and Expert Locators; Web Social Mining

5.Information System Design
Examples: Children and Information Technology; Design Science in the Information Sciences

6.Legal and Ethical Issues
Examples: Art Looting and Trafficking; Cyberspace Law; Piracy in Digital Media

7.Social Life of the Cultural Record
Examples: Open-Access Scholarship and Publishing; Politics of Representation in Museums; Sociology of Reading; Censorship and Content Regulation of the Internet

8.Social Relations in Information Technology
Examples: Collaborative Information Retrieval; Computer-Supported Cooperative Work

9.Social Studies of Information
Examples: Cultural Memory; Digital Divide; Information Policy—European Union; Information Technology Adoption; Organizational Learning; Social Informatics; Social Networks and Information Transfer

Figure 3: Main categories under "Research Specialties" with example entries

I believe that it can be seen from the lists in Figures 2 and 3 how much overlap there is in coverage between the encyclopedia and the iSchools' curricula and faculty research interests.

6. CONCLUSION

The *Encyclopedia of Library and Information Sciences,* Third Edition, not only represents a blueprint laying out the nature and character of the information disciplines, but it also represents the fulfillment of that vision in the recent publication, in seven volumes, of the encyclopedia. It is to be hoped that this reference work will constitute an excellent representation of the i-disciplines, as well as play a role in supporting the further development of those disciplines and their iSchools.

7. REFERENCES

[1] Bates, M.J.; Maack, M.N., Eds. 2010. Encyclopedia of Library and Information Sciences, 3rd Ed., CRC Press.
[2] Bates, M.J. 1999. The invisible substrate of information science. J. Am. Soc. Inform. Sci. 50, 12 (1999), 1043-1050.
[3] Bates, M.J. 2007. Defining the information disciplines in encyclopedia development. Inform. Res., 12, 4 paper colis29. http://InformationR.net/ir/12-4/colis/colis29.html

A Conceptual Model for Scholarly Research Activity

Agiatis Benardou
Digital Curation Unit-IMIS, *Athena* Research Centre
a.benardou@dcu.gr

Panos Constantopoulos
Digital Curation Unit-IMIS, *Athena* Research Centre
& Department of Informatics, Athens University of Economics and Business
p.constantopoulos@dcu.gr

Costis Dallas
Faculty of Information, University of Toronto &
Digital Curation Unit-IMIS, *Athena* Research Centre
& Communication, Media and Culture Department, Panteion University
costis.dallas@utoronto.ca

Dimitris Gavrilis
Digital Curation Unit, Athena Research Centre
d.gavrilis@dcu.gr

ABSTRACT
This paper presents a conceptual model for scholarly research activity, developed as part of the conceptual modelling work within the "Preparing DARIAH" European e-Infrastructures project. It is inspired by cultural-historical activity theory, and is expressed in terms of the CIDOC Conceptual Reference Model, extending its notion of activity so as to also account, apart from historical practice, for scholarly research planning. It is intended as a framework for structuring and analyzing the results of empirical research on scholarly practice and information requirements, encompassing the full research lifecycle of information work and involving both primary evidence and scholarly objects; also, as a framework for producing clear and pertinent information requirements, and specifications of digital infrastructures, tools and services for scholarly research. We plan to use the model to tag interview transcripts from an empirical study on scholarly information work, and thus validate its soundness and fitness for purpose.

Topics
Information seeking and use, Ontologies, Research methods, Scholarly and scientific communication, Digital humanities.

Keywords
Scholarly information behaviour, Conceptual modelling, Digital curation, Activity theory, CIDOC CRM, Cyberscholarship.

1. INTRODUCTION
Research in the arts and humanities relies increasingly on the ability of scholars to discover, appraise, aggregate, organise and use effectively an expanding mass of digital scholarly resources, ranging from primary data and documentary evidence to unpublished and published research, reference works and terminological resources. Current investments in institutional and thematic research e-repositories and digital libraries, and emerging plans for comprehensive digital infrastructures to support scholarly research [1-3] exploiting, among others, the promise of grid technologies, introduce a pressing need to establish a robust conceptual framework for scholarly research information requirements, based on an evidence-based specification of user needs in present and anticipated research work, which will ensure the current and future fitness for purpose of planned systems, applications and standards (metadata, process etc.).

This paper reports on work conducted in the context of *Preparing DARIAH: Preparing for the construction of the Digital Research Infrastructure for the Arts and Humanities*, a collaborative project co-funded by the ESFRI *e-Infrastructures* programme, aiming at providing the foundations (strategic, financial, legal, technological and conceptual) for the timely design and construction of the digital infrastructure requisite for scholarly research in the arts, humanities and cultural heritage in Europe [4]. The Digital Curation Unit-IMIS, *Athena* Research Centre is currently engaging in a two-pronged research programme within the conceptual modelling work-package of DARIAH, consisting: a) of an empirical study of scholarly work, based on the elicitation, transcription, conceptual encoding and interpretation of open-questionnaire interviews with humanities scholars [5], and b) of the formulation of a scholarly research activity model, based on an event-centric modelling approach, and intended to be useful for the formalisation of the analysis of the results of the empirical study.

This paper focuses on the scholarly research activity model developed as part of this project. A summary of earlier work, a definition of research rationale and of the approach adopted, a detailed presentation of the model, and a brief discussion of its utility and planned work are presented below.

2. EARLIER WORK

Earlier studies of research activity, touching upon information practices relevant to this study, range from ethnographies and theoretical syntheses from the field of social studies of science [6-9], which concern mostly the natural and pure sciences, to empirical studies from the field of human information behaviour (HIB) [10], often focusing in humanistic disciplines such as history [11-15] and art history [16-18], and on interdisciplinary research typical in the humanities [19]. A comprehensive overview of concepts, issues, practices and problems related to "scholarship in the digital age", providing a broad framework for conceptualising the relationship between disciplinary practices in the humanities, documents and data, and technological infrastructures and tools, is provided by Borgman [20].

Conceptualisations of scholarly activity, in the form of schematic models and classifications, evolved in tandem with empirical research on information behaviour. Ellis, based on a grounded theory analysis of research practice across the natural, social and human sciences (including economic and social historians, archaeologists, prehistorians and English literature scholars), proposed a classification of research activities composed of six processes, common across disciplines: *starting, chaining, browsing, differentiating, monitoring* and *extracting* [21], to which Meho and Tibbo later added three further processes: *accessing, networking* and *verifying* [22].

The notion of "scholarly primitives" was introduced by Unsworth in 2000, in the context of the then emerging digital humanities field, with reference to the information processes employed by literary scholars [23]. The related concept of "research method" was the focus of the AHRC Methods Network in the UK, which has developed a thesaurus of ICT research methods typically employed by researchers ("Methods taxonomy"), and also a series of brief reports on the needs and plausible scenarios for the current and future scholarly use of ICT in fields such as history, art history, and archaeology [24-26].

Brockman and associates presented a broadly based conceptual framework of the information nature of scholarly work, accounting for processes of *reading, collaborative networking, researching and searching,* and *ways of writing*, and emphasizing the differences in information work in the humanities *vis-à-vis* other disciplines [27]. On the other hand, a study aimed at defining appropriate infrastructures and services at the Minnesota University Libraries [28], was based on organising "scholarly primitives" into four groups of scholarly information activities (*discover, gather, create,* and *share*). Most recently, Palmer and associates defined five broader "scholarly activities": *searching, collecting, reading, writing, collaborating*. These, as well as a bucket of "cross-cutting primitives" are further refined to a more detailed list of twenty granular "scholarly primitives", of which *browsing, collecting, re-reading, assembling, consulting* and *notetaking* were found to be particularly common in the humanities, while *chaining, accessing, assessing, disseminating* and *networking* were seen as equally applicable to all disciplines [29].

Complementary conceptualisations, mostly focussing on *information seeking* behaviour, have emerged from LIS, focussing on the motivations and sequential actions of researchers as they seek information (from the stage of initiation, to selection/exploration, formulation, collection and, finally, presentation) [30], on the process, factors and mechanisms affecting information seeking, including notions of context of information need, psychological, role-related and interpersonal factors, social learning, and search strategies [31,32]; on information seeking as problem solving, employing notions of goal-directed behaviour, resolving an "anomalous state of knowledge", or reducing uncertainty [33-36], or, on everyday life information seeking as sense making [37]. An overview of information seeking behaviour work up to 2002 is provided by Case [38]; an integrative model of information behaviour synthesising problem solving, sense making, information foraging and modular thinking approaches, was later proposed by Spink and Cole [39].

These models view information behaviour primarily as process; consequently, the world of information objects, data, and documents, remains in them as a rule implicit. Exceptions include Ingwersen's cognitive model, informed by information retrieval system design, and viewing the information seeking universe as a set of cognitive transformation and interactive communication relationships between *information objects*, an individual user's *cognitive space* (and social-organisational environment), and an *IR system* [40]. Also, Saracevic' model of stratified interaction (distinguishing between *surface*, *cognitive* and *situational* layers) conceptualises information use via a sequence of interactions between *environment, situation, user knowledge etc., query characteristics, interface, computational resources,* and *informational resources* [41]. Finally, an implicit conceptual model is provided by the Minnesota study in the form of an extensive graphical diagram "track[ing] relationships between a) *primitives*, b) common *tasks*, c) support from *data*, and d) potential *tools* and *services* that would address scholars' needs" (p. 47, our emphasis) [28].

3. RESEARCH RATIONALE AND APPROACH

As presented above, there is already a multitude of conceptualisations of scholarly information activity, or information behaviour in general, including useful classifications of specific scholarly activities and research methods, and macro-analysis models of human information behaviour, accounting for motives, goals and research strategies, and/or for sequential structure of information practices. These conceptualisations:

1. Are concerned predominantly with practices of information seeking, or searching, rather than on the whole lifecycle of scholarly information use, including curation activities (structuring, annotation, processing) typical of actual scholarly practice.
2. Focus mostly on the use of scholarly objects – research works, publications – from a library service perspective, and only implicitly on primary evidence (data, documentary sources) and hybrid, secondary archives, which, in fields such as history, art history and archaeology [42-44,27], constitute a central object of scholarly engagement in the research process.
3. Privilege process over object modelling, and thus account predominantly for factors (psychological, socio-technical, environmental) governing human information behaviour and relations thereof, and/or sequential/procedural structure of the information seeking process, rather than for the universe of entities (material, informational or conceptual) involved in information work.
4. Delineate collectively a broad domain of diverse entities of interest in the research process (such as "primitives", research activities, methods, goals, motives, strategies, data and computational objects of various kinds), but such entities are only informally or extensionally defined in individual studies; in fact, there is no single model representing formally entities involved in the research process and relations between them.
5. Have mostly the status of explanatory schematisations, linking together specific factors, theoretical perspectives and implicit relationships, rather than of formal conceptual models amenable to operationalisation through specific bindings to data structures and procedural logic.

Our objective is, thus, to establish a conceptually sound, pertinent with regard to actual scholarly practice, and elegant model of scholarly research activity, encompassing both "object" (structure) and "process/practice" (functional) perspectives, and amenable to operationalisation as a tool for:

- structuring and analysing the outcomes of evidence-based research on scholarly practice and requirements, and
- producing clear and pertinent information requirements, and specifications of architecture, tools and services for scholarly research in a digital environment.

Our approach is inspired by Leont'ev's cultural-historical activity theory, used as a useful framework in diverse fields, including developmental psychology, the study of organisations, work and ergonomics, social aspects of technology, Human-Computer Interaction, and digital curation [45-48]. Its key concept is *activity,* understood as "purposeful interaction of a subject with the world"; an activity is always directed toward some *object*, a physical or conceptual entity (or entities). This object embodies, also, the fulfilment of some objective or *motive,* which in turn is intended to meet a specific *need* of the subject of the activity. Activity systems are composed as a hierarchy of *activities,* constituted by conscious *actions,* which in turn are constituted by subconscious *operations;* actions are designed to meet hierarchically structured goals. Subjects can be *individuals*, but also *communities* with shared needs and motives. Purposeful *interaction* between subjects and objects takes place by means of *tool mediation,* whereby tools include not just physical things, but also procedures, computer programs, languages and signs [47].

Modelling the research process is informed by the preliminary findings of our empirical study of scholarly practice, as well as literature cited above. The study was carried out in the form of semi-structured conversational interviews, so far with 23 European arts and humanities scholars who may be classified as mainstream users of digital resources and tools [5]; a second round of interviews, currently under way, involves scholars who could be classified as innovators or early adopters *vis-à-vis* ICT use. Interviews were recorded, transcribed in machine-readable form, and tagged, using an initial corpus of tags derived from information behaviour and research methods literature [21,23,28,29], and adding further tags when dictated by the conceptual content of the source material. Candidate concepts, such as activity, procedure, method, tool, and information object, were abstracted from segmented interview transcripts and associated tags. Actual modelling of the scholarly research process was based on established conceptualisations of activity, such as the concept of activity in the CIDOC CRM cultural documentation ontology [49], and the concept of process in enterprise information models [50,51].

4. CONCEPTUAL MODEL

Consistent with the activity theory framework presented above, scholarly research is here understood as a *purposeful process*, carried out by *actors,* individuals or groups, according to specific *methods*. Research processes usually are complex, consisting of simpler *tasks,* which may be carried out in parallel or in series. Each task may be further analyzed, until we arrive at elementary tasks. The detailed structure of the research process, and way of working for each step, are specified by a corresponding *procedure*. Procedures have a normative character and convey what is believed by a community of practitioners to be good practice at any given time.

A research process can be considered as an enactment of the corresponding procedure, carried out at a specific place and time by a specific individual or group. To capture the purpose of a particular instance of research work and of the successive steps of the process, a structured representation of goals is needed. At the highest level, we wish to express what the research is after, and why, and to determine, probably in general terms, its felicity conditions. As we proceed to the tasks and sub-tasks of the research process, goals become more specific and can be associated with the performance of services designed to support the respective tasks.

Not being conceived as merely a structured set of events (*what, where, when was done?*), but rather being characterized by subject (*who did it?*), method (*how was it done?*) and purpose (*why was it done?*), allows the research process to be considered as a special case of the notion of *activity* of the CIDOC CRM ontology (formally, a subclass of the Activity class) [49]. There is, however, one significant teleological difference between that model and the one we are

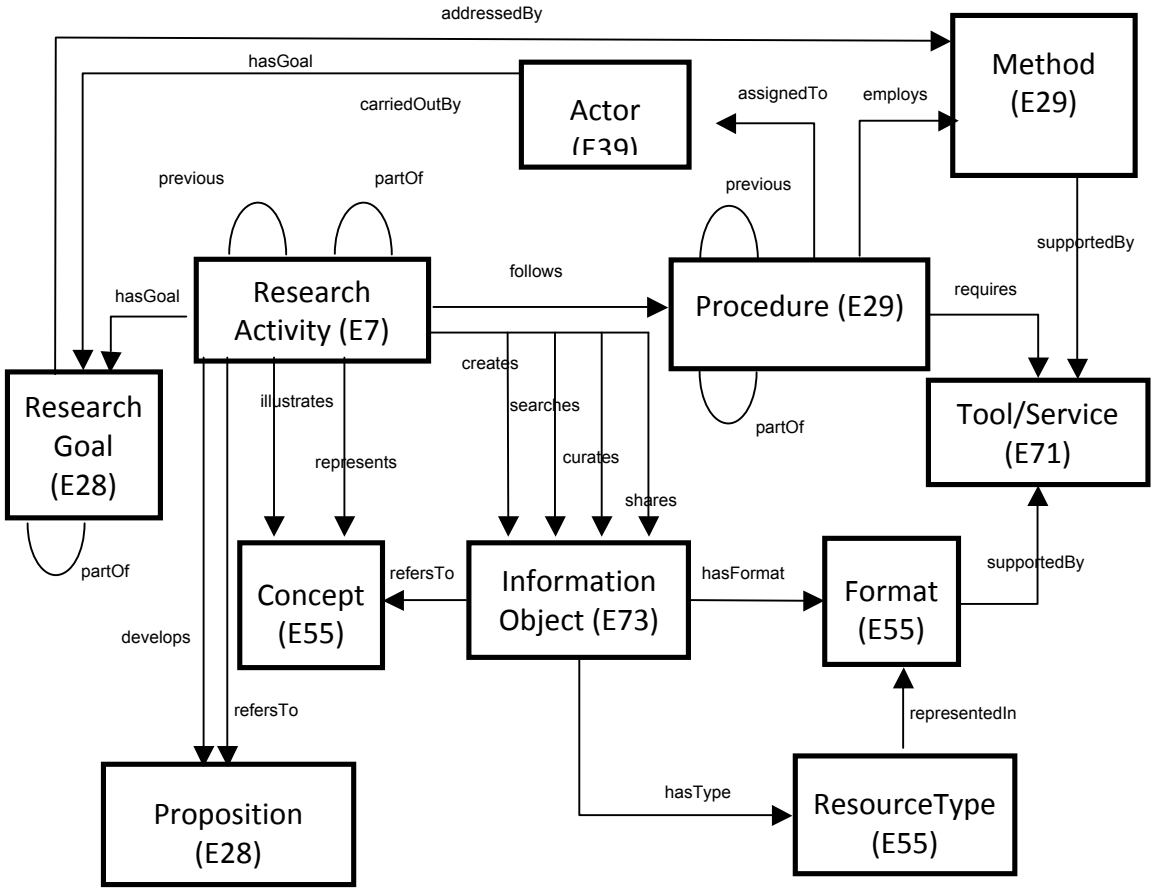


Fig. 1: Scholarly research activity

proposing here: the activity model of CIDOC CRM is meant for historical and documentary purposes. Our model of the research process is intended to facilitate the design and development of information repositories and services in digital infrastructures that support research in the arts and humanities. To this end the model should allow to represent the details of a particular research activity both at the level of planning (*how it should be done*) and of actual execution. This dictates the distinction between process and procedure while maintaining corresponding (though not necessarily isomorphic) descriptions of the two. This duality is commonly encountered in conceptual models of task-oriented systems, such as enterprise information systems [50,51].

A research activity involves a range of objects of different kinds: physical objects (natural or artificial), conceptual objects and information objects. *Physical objects* are those found, examined, stored, etc., or those used as tools in the course of the research process. *Conceptual objects* include concepts created, represented and illustrated, and logical propositions formulated, supported, countered, proved, disproved or refuted. The *information objects*, finally, are a special class of conceptual objects with corresponding physical information carriers, which refer to and represent physical and conceptual objects, and which are created, searched, shared, or even curated as part of research processes. Each of these categories actually gives rise to a representational facet of autonomous interest, related to the others through the relevant research activity: the information objects are the contents of digital repositories; the physical objects are the original domain material; and the conceptual objects are the content of scientific theories. Our work focuses on the interplay between conceptual and information objects, so physical objects are not considered here any further.

We now turn to Fig.1 for a schematic presentation of the conceptual model of scholarly research activity. Next to the name of each entity in parentheses we note the code of an appropriate CIDOC CRM superclass. By property inheritance, the entities in our model share all the properties assigned to the respective CIDOC CRM superclasses and we decline listing those here, with one exception for illustration purposes.

The entity *Research Activity* is the basic construct for representing research processes. Being a subclass of the CIDOC CRM E7 Activity, this entity is endowed with all the properties describing E2 Temporal Entity, E4 Period and E5 Event (successive classes on the hierarchy path above E7) in addition to those of E7. Of particular interest in our case are: P4 (*has time-span*), P119 (*meets in time with*), P7 (*took place at*), P9 (*consists of*), P11 (*had participant*), P14 (*carried out by)*, P16 (*used specific object*), P17 (*was motivated by*), P20 (*had specific purpose*), P21 (*had general purpose*), P125 (*used object of type*), P134 (*was continued by*). These are not shown in Fig.1 for the sake of clarity. Nevertheless, the prop-

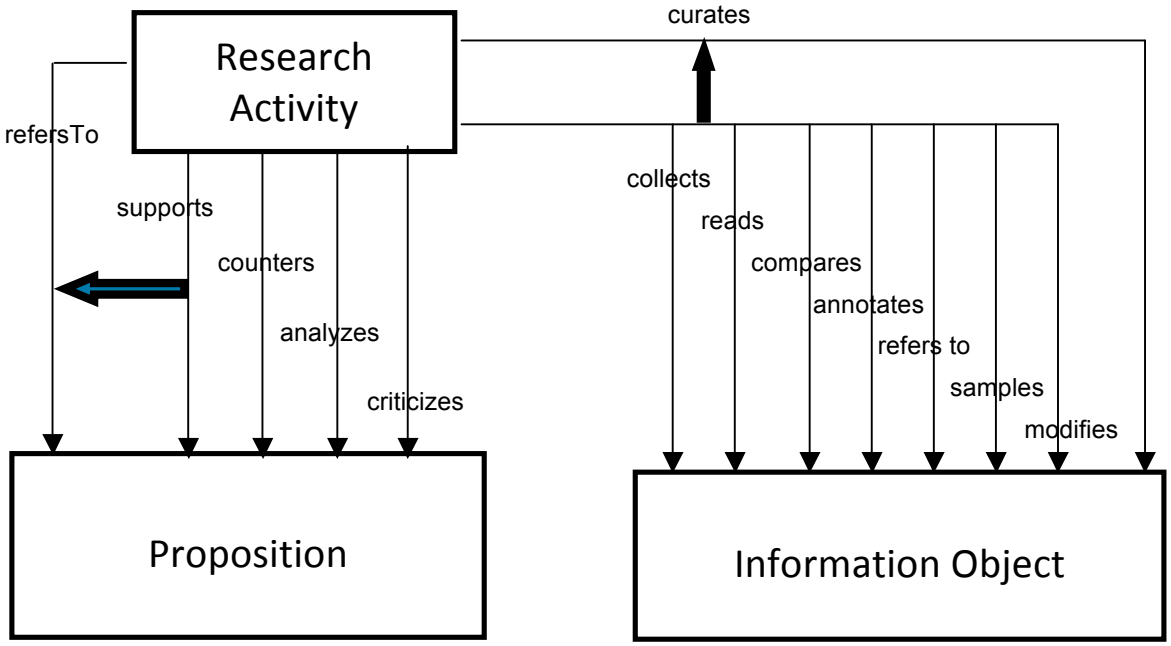


Fig.2: Modelling scholarly primitives

erties *previous* (sub-property of P134) and *partOf* (sub-property of P9) of *Research Activity* are shown in order to stress their prominent role in defining structure. This entity can be used for the documentation of accomplished as well as planned research processes through respective subclasses. Comparisons of corresponding property values allow inferences on the applicability of procedures and the actual use of resources.

The association of *Research Activity* with *Procedure* is the key element for recording normative aspects and planning. Besides having a similar structure with a research activity, a procedure is related to the *Methods* it employs and the *Tools* and software *Services* it requires. The latter can be a selection from the tools and services that support the respective methods. *Methods* may also be identified as useful in addressing particular research goals. The specification of a set of research goals with an appropriate internal organization (usually, but not necessarily, hierarchical) is captured by the entity *Research Goal*.

The *Proposition* entity represents all logical propositions, such as hypotheses formulated, inferences made, arguments raised for or against other propositions. The generation of propositions is represented by the *develops* property of *Research Activity*, while any other kind of reference to them is represented collectively by the *refersTo* property. Propositions refer to concepts and objects. *Concepts* are (or should be) represented and organized in appropriate thesauri. Objects are represented and documented by *Information Objects*. These are *created* in the course of research activities and populate digital repositories through which they can be *searched* and *shared*. Information objects of different types can be represented in specific *formats* which, in turn, require specific software *services* for access and processing. *Services* thus become an important mediator between methods, procedures and information repositories. From a functional perspective, affordances of digital scholarship are embodied in services available. From a teleological and methodological perspective, services evolve to better meet requirements.

Previous has authors have identified various sets of "scholarly primitives" as basic operations that take place in scholarly research and that can be used both to understand how a scholar works and which functions to support when designing tools for scholarly use. The model presented here provides a more general framework, in which scholarly primitives can be interpreted as specific operations on conceptual or information objects. Accordingly, they may be represented as specializations of properties relating *Research Activity* to *Proposition, Concept* and *Information Object.* Studies like those cited above, or like our own empirical study, provide the necessary substantiation on primitives which, together with an elaboration of research goals, enables developing a model of scholarly research processes specific enough to support development of appropriate digital services.

In Fig. 2 we present one possible specialisation of the properties (in RDFS triples style) <*Research Activity, refersTo, Proposition*> and <*Research Activity, curates, Information Object*>. The proposed specializations of *refersTo* are: *supports, counters, analyzes*, and *criticizes*. These are genuine scholarly functions that actually refer to scholarly statements (*Propositions*), which can be mapped onto appropriate annotation functions operating on data objects containing the statements. On the other hand, the *curates* property is specialized into a set of properties that correspond to scholarly primitives that actually involve curating information objects: *collects, reads, compares, annotates, refers, samples*, and *modifies*. Clearly, different sets of primitives can be accommodated by the model in exactly the same way.

5. DISCUSSION

The conceptual model presented here aims to address a press-

ing need as we engage in the empirical elicitation, analysis, abstraction, and formal conceptualisation of information practices and needs in scholarly research. It has been developed as a complement, rather than as an alternative, to prior conceptualisations of information work, including classifications of scholarly activities and methods with which it can be interoperable. It provides clear definitions of constituent entity types and relationships, distinguishing, in particular, the cardinal concept of *Research Activity* from those of *Procedure, Method, Tool/Service*, and elucidating the relationships between *Information Object, Resource Type, Format,* and *Concept*. It also provides, by way of illustration, specialisations of the relationship between *Research Activity* and *Proposition*, and that between *Research Activity* and *Information Object,* mapping to empirically attested operations in research practice.

With the important proviso of the need to maintain a distinction between (empirically attested) process, or activity, and (goal- and task-driven) procedure, the model is fully expressed in terms of the CIDOC Conceptual Reference Model, a mature, internationally recognised standard for cultural information [49]. It is amenable to operationalisation, i.e., as a conceptual schema for the construction and population of a knowledge base with facts regarding information practices of humanities researchers, and corroborating documentary evidence, such as interview transcripts produced by our empirical study [5]. In the next stage of our conceptual modelling work in the DARIAH project, we plan to tag systematically all interview transcripts in terms of the model, and thus provide, in addition, a mechanism of validation of its soundness and fitness for purpose.

At the same time, the model is meant to act as a descriptive framework for better discovery, summarisation and understanding of relationships between specific scholarly activities, research goals, information objects, methods, and tools/services at the instance level. It may, therefore, be useful as a conceptual structure – or information architecture – for better communication among stakeholders (such as policy makers, archivists, repository managers, technologists and scholars) and institutions involved in the specification of requirements and affordances of digital repositories, services and tools intended to support scholarly research work.

6. ACKNOWLEDGMENTS

Authors acknowledge the financial support of the European Commission *e-Infrastructures* programme to research presented here. They also wish to thank their DCU colleagues, and their partners in *Preparing DARIAH* (workpackages 7 and 8) for useful suggestions and constructive criticism.

REFERENCES

[1] ACLS, *Our cultural commonwealth: the report of the American Council of Learned Societies Commission on cyberinfrastructure for the humanities and social sciences*, American Council of Learned Societies, 2006. http://www3.isrl.illinois.edu/~unsworth/sdl.html

[2] T. Blanke and S. Dunn, "The arts and humanities e-Science initiative in the UK," *Second IEEE International Conference on e-Science and Grid Computing, 2006. e-Science'06*, 2006, pp. 136–136

[3] G. Crane, A. Babeu, and D. Bamman, "eScience and the humanities," *International Journal on Digital Libraries*, vol. 7, 2007, pp. 117–122

[4] P. Constantopoulos, C. Dallas, P. Doorn, D. Gavrilis, A. Gros, and G. Stylianou, "Preparing DARIAH," *Proceedings of the International Conference on Virtual Systems and MultiMedia (VSMM08)*, Nicosia, Cyprus: 2008. http://www.dcu.gr/dcu/Documents/documents/preparing-dariah/en/attachment

[5] A. Benardou, P. Constantopoulos, C. Dallas, and D. Gavrilis, "Understanding the information requirements of arts and humanities scholarship: implications for digital curation," *International Journal of Digital Curation,* 5, 1, forthcoming

[6] B. Latour, *Science in action: How to follow scientists and engineers through society*, Harvard University Press, 1987

[7] G.C. Bowker, *Memory practices in the sciences*, MIT Press Cambridge, MA, 2005

[8] B. Latour and S. Woolgar, *Laboratory life: The construction of scientific facts*, Princeton University Press, 1986

[9] K. Knorr-Cetina, *Epistemic cultures: How the sciences make knowledge*, Harvard University Press, 1999

[10] S. Stone, "Humanities scholars: information needs and uses," *Journal of Documentation*, vol. 38, 1982, pp. 292-313

[11] M.S. Dalton and L. Charnigo, "Historians and their information sources," *College and Research Libraries*, vol. 65, 2004, pp. 400-426

[12] R. Delgadillo and B.P. Lynch, "Future historians: their quest for information," *College and Research Libraries*, vol. 60, 1999, pp. 245-260

[13] W. Duff, B. Craig, and J. Cherry, "Historians Use of Archival Sources: Promises and Pitfalls of the Digital Age," *The Public Historian*, vol. 26, 2004, pp. 7-22

[14] H.R. Tibbo, "Primarily history: historians and the search for primary source materials," *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, Portland, OR: ACM, 2002, pp. 1-10

[15] H.R. Tibbo, "Primarily History in America: how US historians search for primary materials at the dawn of the digital age," *American Archivist*, vol. 66, 2003, pp. 9-50

[16] W.O. Beeman, "Stalking the art historian," *Work and Technology in Higher Education: The Social Construction of Academic Computing*, 1994, p. 89

[17] W.S. Hemmig, "The information-seeking behavior of visual artists: a literature review," *Journal of Documentation [Bradford]*, vol. 64, 2008, pp. 343-362

[18] L. Odum, "The uses of archival materials by art historians," MSLS Thesis. University of North Carolina at Chapel Hill, 1998

[19] C.L. Palmer and L.J. Neumann, "The information work of interdisciplinary humanities scholars: Exploration and translation," *The Library Quarterly*, vol. 72, 2002, pp. 85-117

[20] C.L. Borgman, *Scholarship in the digital age: information, infrastructure, and the Internet*, Cambridge, MA; London: MIT Press, 2007

[21] D. Ellis, "Modeling the information-seeking patterns of academic researchers: A grounded theory approach," *The Library Quarterly*, vol. 63, 1993, pp. 469-486

[22] L.I. Meho and H.R. Tibbo, "Modeling the information-seeking behavior of social scientists: Ellis' study revisited," *Journal of the American Society for Information Science and Technology*, vol. 54, 2003, pp. 570-587

[23] J. Unsworth, "Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this?" King's College, London: 2000. http://www3.isrl.illinois.edu/~unsworth/Kings.500/primitives.html

[24] N. Grindley, "What's in the art historian's toolkit? A Methods Network working paper," Sep. 2006. http://www.methodsnetwork.ac.uk/redist/pdf/wkp01.pdf

[25] N. Grindley, "Digital tools and methods for historical research: a Methods Network working paper," Dec. 2006. http://www.methodsnetwork.ac.uk/redist/pdf/wkp04.pdf

[26] N. Grindley, "Digital tools for archaeology: a Methods Network working paper," Feb. 2007. http://www.methodsnetwork.ac.uk/redist/pdf/wkp06.pdf

[27] W.S. Brockman, L. Neumann, C.L. Palmer, and T.J. Tidline, "Scholarly Work in the Humanities and the Evolving Information Environment," Dec. 2002. http://www.diglib.org/pubs/dlf095/index.htm

[28] University of Minnesota Libraries, *A multi-dimensional framework for academic support: final report*, Minneapolis: University of Minnesota Libraries, 2006. http://purl.umn.edu/5540

[29] C.L. Palmer, L.C. Teffeau, and C.M. Pirmann, *Scholarly Information Practices in the Online Environment*, Dublin, Ohio: OCLC, 2009. http://0-www.oclc.org.millennium.mohave.edu/programs/publications/reports/2009-02.pdf

[30] C.C. Kuhlthau, "Inside the search process: information seeking from the user's perspective," *Journal of the American Society for Information Science*, vol. 42, 1991, pp. 361-371

[31] T.D. Wilson and C. Walsh, *Information behaviour: an interdisciplinary perspective*, Sheffield: University of Sheffield, Department of Information Studies, 1996

[32] T.D. Wilson, "On user studies and information needs," *Journal of Documentation*, vol. 37, 1981, pp. 3-15

[33] A. Foster, "A nonlinear model of information-seeking behavior," *Journal of the American society for information science and technology*, vol. 55, 2004, pp. 228-237

[34] T.D. Wilson, "Models in information behaviour research," *Journal of Documentation*, vol. 55, 1999, pp. 249-270

[35] N. Belkin, "Anomalous states of knowledge as a basis for information retrieval," *Canadian Journal of Information Science*, vol. 5, 1980, pp. 133-134

[36] C.C. Kuhlthau, *Seeking meaning: a process approach to library and information services*, Norwood, N.J.: Ablex, 1993

[37] B. Dervin, "On studying information seeking methodologically: the implications of connecting metatheory to method," *Information Processing & Management*, vol. 35, 1999, pp. 727-750

[38] D.O. Case, *Looking for information: a survey of research on information seeking, needs, and behavior*, San Diego, CA: Academic Press, 2002

[39] A. Spink and C. Cole, "Human information behavior: Integrating diverse approaches and information use," *Journal of the American Society for Information Science and Technology*, vol. 57, 2006, pp. 25-35

[40] P. Ingwersen, "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory," *Journal of Documentation*, vol. 52, 1996, pp. 3-50

[41] T. Saracevic, "Modeling interaction in information retrieval (IR): a review and proposal," S. Hardin, ed., Silver Spring, MD: American Society for Information Science, 1996, pp. 3-9

[42] C. Dallas, "Archaeological knowledge, virtual exhibitions and the social construction of meaning," *Virtual museums and archaeology: the contribution of the Italian National Research Council*, P. Moscati, ed., Roma: Insegna del Giglio, 2007, pp. 31-64

[43] C. Dallas, "Humanistic research, information resources and electronic communication," *Electronic Communication and Research in Europe*, J. Meadows and H. Boecker, eds., Luxembourg: European Commission, 1999, pp. 209-239

[44] K. Cohen, J. Elkins, M.A. Lavin, N. Macko, G. Schwartz, S.L. Siegfried, and B.M. Stafford, "Digital Culture and the Practices of Art and Art History," *The Art Bulletin*, vol. 79, Jun. 1997, pp. 187-216

[45] A.N. Leont'ev, "Activity, consciousness and personality," May. 2007. http://lchc.ucsd.edu/MCA/Paper/leontev/index.html

[46] Y. Engeström, "Activity theory as a framework for analyzing and redesigning work," *Ergonomics*, vol. 43, 2000, pp. 960-974

[47] V. Kaptelinin and B.A. Nardi, *Acting with technology: activity theory and interaction design*, Cambridge, MA & London: MIT Press, 2007

[48] C. Dallas, "An agency-oriented approach to digital curation theory and practice," *The International Cultural Heritage Informatics Meeting Proceedings,* J. Trant and D. Bearman, eds., Toronto: Archives & Museum Informatics, 2007. http://www.archimuse.com/ichim07/papers/dallas/dallas.html

[49] N. Crofts, M. Doerr, T. Gill, S. Stead, and M. Stiff, eds., Definition of the CIDOC Conceptual Reference Model (version 5.0.1), ICOM/CIDOC CRM Special Interest Group, 2009. http://cidoc.ics.forth.gr

[50] D.C. Hay, *Data models: conventions of thought,* Dorset House, 1996

[51] J. Dietz, *Enterprise ontology – theory and methodology,* Springer Verlag, 2006

Backstage or Front Stage with YouTube

Timothy D Bowman
Indiana University
School of Library and Information Science
1320 East 10th Street, LI 011, Bloomington, IN 47405-3907
812.855.2018
tdbowman@indiana.edu

ABSTRACT
This paper explores backstage behavior in videos found by searching for "drinking and puking" on YouTube. A small sample of 10 videos was critiqued using the definition of backstage language behavior found in Goffman's *Presentation of Self in Everyday Life.* The question examined is: Is there a blurring of the boundaries between front stage and backstage behavior in videos posted to YouTube? Three possibilities emerge from the research relating to boundary establishment in this mediation of social interaction by technology.

Categories and Subject Descriptors
H5.m. **Information interfaces and presentation** (e.g., HCI): Theory and methods.

General Terms
Theory

Keywords
Goffman, front stage, backstage, YouTube, video, social interaction, technology mediation

1. INTRODUCTION
The Internet has had a tremendous impact on the way we communicate in our daily lives. It has changed the way we interact socially and opened up a variety of avenues for expressing ourselves individually. Tools on the Internet such as blogs, personal home pages, and social network sites are among a variety of technologies that utilize the computer and Internet to facilitate new types of social interaction, community building, and communication. As one writer states, "technology has provided us with new sites of empirical experience and it has re-configured the complex ties that bind the social and the cognitive worlds" [3; p. 55]. This expansion of our social environment has led the author to question the ways in which technology mediates social interaction. An example where technology can be seen mediating interaction can be found on the YouTube web site (www.youtube.com).

YouTube provides us with a perfect example for examining the phenomenon of social interaction mediated by technology. There are a range of social theories that might be of interest when analyzing this phenomenon. Of particular interest is Goffman's dramaturgical theory. Among the many concepts involved in the dramaturgical theory, the most interesting is the concept of region and region behavior. In this preliminary study, Goffman's framework is used to examine a small selection of YouTube videos and critique them within the context of his definition of backstage behavior. The question of interest in this preliminary work can be stated as: Is there a blurring of the boundaries between front stage and backstage behavior in videos posted to YouTube?

2. YOUTUBE
YouTube is a web site dedicated to the distribution of online videos. The site currently has 55 million unique users each month and has the 8th largest audience on the Internet [18]. YouTube brought video sharing into the mainstream by providing the ability for videographers to easily upload videos and tag videos with keywords. A visitor to YouTube can browse video categories, user-created channels, communities, or simply search by keyword. Visitors can create profiles, join live video streams, leave comments on each video, or rate videos. The site also offers a "related content" feature that provides visitors with a list of videos with similar keywords and titles.

By utilizing the Adobe Flash video player, YouTube presents videos in a single format which simplifies visitor requirement. Through the use of this video streaming technology, YouTube establishes a single media-player platform across the entire site. YouTube also allows for simple video sharing by providing html tags on each video page allowing visitors to copy and paste HTML code into other web sites such as MySpace, Facebook, or any other site that allows this copy and paste behavior.

Videos are uploaded to the YouTube website, waiting to deliver their content to any visitors who happen across them. These YouTube videos present us with a multitude of actors, teams, and performances. In this study, Goffman's theory, specifically backstage behavior, has been used to examine 10 of these videos.

3. GOFFMAN
In [7], Erving Goffman provides a framework for examining social interactions in everyday experience. Dramaturgical concepts are used to interpret performed roles and deduce social meaning by examining an individual's role during an interaction. A performance, in this framework, is defined as a setting in which an individual (actor) performs a distinct role given for the benefit of an observer (audience). The impressions the actors give and give off during a performance are defined by Goffman as sign vehicles.

During a performance, the actor or actors are considered a team. Similar beliefs and behaviors are emphasized in the performance, signifying to the audience that the actors are part of the same team. Any disagreements between team members are discussed away from the audience. An impression is maintained by the team members at all times while in front of the audience. One of the primary motivating factors for establishing and maintaining a consistent impression is the avoidance of embarrassment [12]. Because of this behavior, a clear boundary is established between the audience and the actor/team. The audience can also be considered a team, acting in accordance with other audience members in response to the presentation before them. [7] considers this interaction among the two teams a dramatic interaction, a give and take between the actor(s) and audience that is central to avoiding embarrassment.

This boundary between audience and team is defined as a region. Goffman divides regions into areas of front stage, backstage, and the outside. Front stage behavior takes place before an audience; the place where actors perform for the audience while meeting standards and expectations of social performance. When the actors are at a pause from performing for the audience and are amongst fellow team members separated from the audience, they are considered in the backstage region. Lastly, the area that is not considered part of the front or backstage but separate from the performance is defined as the outside area. By dividing interaction into these three regions, Goffman has given us distinct boundaries in which teams, actors and audience members establish rules and regulations for proper interaction behavior. Access between front stage and backstage is generally controlled in order to prevent audiences from coming backstage or to prevent audiences from seeing a performance that was not given for them.

Goffman's theory provides a framework that allows us to explain our social interaction. We constantly create stages in our day-to-day lives in which we act according to social norms and follow behaviors consistent with our situation to avoid embarassment. Each situation affords us a new constraint, shifting from front stage to backstage, audience to performer. We are also presented with moments as outsiders, in which we come upon a performance that was unintended for our consumption.

Videos on YouTube present another case in which we are presented with a performance. Videos are placed on the Internet, waiting to deliver their performance to a visitor who only has to push play. Although Goffman's theory focuses on face-to-face interaction, other research has shown that it can be a valuable theory when examining online sources.

4. LITERATURE REVIEW

A selective review of the literature reveals a variety of uses for Goffman's front stage and backstage region definitions. In these research articles, a rigid boundary was shown to exist between front stage and backstage behavior in a variety of social settings. This is significant because it provides relevance for this preliminary study and shows that Goffman's dramaturgical framework can provide insight into social interaction behaviors.

In social setting devoid of technology, we find that boundaries between front stage and backstage regions exist. While examining a support group for pregnant women, [16] found distinct separation between front stage, backstage, and what was termed "back-backstage" communication behavior. Communication behavior was recorded in "play-group" meetings, "night-out" meetings, and private discussions. Front stage behavior during play-group meetings revealed discussions of health care, doctors, appointments and tests and procedures. These discussions were limited to formal conversation and did not involve backstage behavior. Backstage communication behavior was observed during night-out gatherings. During night-out gatherings, the women were without children or husbands and discussed items they would not share during the more formal, play-group meetings. The writer noted that "without their children or husbands present, the women were able to discuss topics otherwise not discussed among their children or husbands" [16; pg. 459]. In private discussion among women, it was found that taboo discussions were limited to what Tardy termed "back-backstage" [16; pg. 462]. These discussions were held in strict privacy and involved issues such as sexual relations and sexual diseases.

In studies of social communications mediated by technology, we also see evidence of distinct front stage and backstage behavior. A researcher examined an organic online learning community (OOLC) and determined that language and pseudonymity were two important aspects for defining back regions [11]. Results found that using community-specific language in an online community allowed participants to include members of their own community while excluding outsiders. Observations also found that using pseudonyms to identify oneself in this online community provided for users a separation between front stage and backstage behavior. This allowed users to "reduce or eliminate the consequences of practicing FR [front region] performances, criticizing the FR and engaging in 'inappropriate' banter" [11; p. 321]. By providing this separation of front regions and back regions, the OOLC back region became "a sanctuary of sorts for taking academic and social risks, one where potential consequences to offline reputations are few" [11; p. 322].

Personal home pages were examined and [14] describe the occurrence of indirect and direct modes of self-presentation. Indirect modes of self-presentation were defined as whatever information was posted about the person on the page. These could include names, descriptions, or images of the person. Linking behavior was also discussed and related to Goffman's idea of team performance. Direct modes of self-presentation were defined as ways in which a person highlights aspects of self, while at the same time omitting other aspects that might seem inappropriate or secret. In regards to backstage behavior, the authors note that "the only way a visitor might access backstage information would be if someone on a performance team presented contradictory or unflattering information in his or her link" [14; pg. 9].

While examining blogging and blogging behavior, [8] found that bloggers portray an idealized version of themselves through their blogging practices. The findings state that blogs "provide a way to understand ourselves by inscribing ourselves into a new type of text" [8; p. 65]. She also believes a blog can loosely be defined as a front stage presentation of self; "The blog is a case where the human personal front is mediated by the technology to create a front hybrid, with new mutabilities and new durabilities" [8; p. 66].

In these examples, Goffman's dramaturgical theory has been shown to provide a distinct boundary between front stage and backstage behavior in a variety of contexts. In this study, Goffman's definition of backstage behavior has been used to examine these videos.

5. METHODS

For this preliminary study, a small sample of 10 YouTube videos was critiqued using Goffman's definition of backstage language behavior. Because much of Goffman's work focuses on the idea of embarrassment and the avoidance of shame, videos were sought in which the behavior presented might afford shame or embarrassment to the actors if viewed by unintended audiences. After a preliminary review of videos found on YouTube, videos with vomiting behavior were selected because of the taboo associated with expulsion of human biological waste. Each video was found by searching for "drinking and puking" within the YouTube web site. The recommendation section, found on each video's page, was used to select the next video observed. This was not a random sample; subsequent videos were chosen based on title and keywords presented in the recommendation section. A random selection was initially tried, but videos unrelated to drinking or puking behavior were consistently chosen.

It is important to note that sampling Internet sources can be problematic. YouTube videos can be short-lived; they may be removed at any time by the user who uploaded them or may be taken down by YouTube if they are found to offensive or in violation of copyright. Also, there is no way of knowing how many videos exist on YouTube at a given moment.

To analyze the videos, Goffman's definition of backstage behavior was used to code a subset of YouTube videos. Goffman listed the following criteria for defining backstage language: "The backstage language consists of reciprocal first-naming, co-operative decision-making, profanity, open sexual remarks, elaborate griping, smoking, rough informal dress, "sloppy" sitting and standing posture, use of dialect or sub-standard speech, mumbling and shouting, playful aggressivity and "kidding", inconsiderateness for the other in minor but potentially symbolic acts, minor physical self-involvements such as humming, whistling, chewing, nibbling, belching, and flatulence" (pg. 128). The category "inconsiderateness for the other in minor but potentially symbolic acts" was not used for analysis because the category is not clearly definable. Although this is not an exhaustive list of backstage behavior, for this preliminary examination into this dramaturgical theory it was determined to be a good first step in classifying behavior presented in the videos.

The focus of this analysis was on videos that contained the terms "drinking" and/or "puking" in title or tags as defined by the owner. Goffman stated that "another area is suggested by the very widespread tendency in our society to give performers control over the place in which they attend to what are called biological needs. In our society, defecation involves an individual in activity which is defined as inconsistent with the cleanliness and purity standards expressed in many of our performances" [7; p. 121]. Public show of vomiting behavior can also be placed in this category of biological needs. In our current western society, public vomiting is not an accepted behavior. For this reason, backstage behavior was being portrayed in a front stage manner.

Table 1. Occurrence of backstage behavior per video.

	Videos									
Behavior	[2]	[6]	[10]	[1]	[15]	[13]	[17]	[9]	[5]	[4]
Reciprocal First-Naming	Yes		Yes		Yes		Yes	Yes	Yes	Yes
Co-operative Decision-making	Yes				Yes	Yes	Yes	Yes	Yes	
Profanity	Yes		Yes	Yes	Yes	Yes			Yes	Yes
Open Sexual Remarks									Yes	
Elaborate Griping					Yes					Yes
Smoking								Yes		
Rough Informal Dress	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes	
"Sloppy" Sitting and Standing Posture	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Use of Dialect or Sub-standard Speech	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mumbling and Shouting	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes
Playful Aggressivity and "Kidding"					Yes	Yes			Yes	
Inconsiderateness for the Other in Minor but Potentially Symbolic Acts	Not Utilized									
Minor Physical Self-involvements	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

6. SECTIONS

The results presented below are broken down by each of 13 backstage criteria. In table 1, a simple breakdown of the 10 videos and the occurrence of backstage behaviors are presented. In each video, backstage behaviors were recorded. In 70% of the videos, seven or more categories of backstage behavior were coded as existing.

6.1 First-naming

In seven out of the ten videos analyzed, first name or nick-name use occurred. In [2], we hear the name "Pete" several times referring to the young man who is the focus of the video. They use the name to encourage his drinking behavior, and to subsequently provoke him and antagonize him about his vomiting behavior. In [10], the title and description assigned to the video both contain the name "Vince." The videographer focuses on a young man who is chugging a can of beer. The young man finishes chugging his beer and makes a gagging motion, prompting the audience members around him to turn and look at him. Off camera, several voices are heard acknowledging his accomplishment and using the name "Vince" to refer to the young man.

The third occurrence, found in [15], utilizes the first-naming behavior in the title, keywords, and description of the video. This video focuses on a young man who finishes a glass of whiskey in one drink. In the video, we hear the name "Parsons" utilized to refer to the drunken young man and we hear the name "Jason" in reference to another person in the performance. The fourth instance, found in [5], depicts a young man passed who is subsequently carried by two other men into a bathroom and dropped into a bathtub of water. Several times the passed out man is referred to as "Al" or "Allen." Later in the video, another man refers to the videographer as "Randy." [4], the final video utilizing first name behavior within the video itself, shows four men chugging beers. In the final moments of the video, we hear the name "Mort" in reference to a question we hear posed off camera.

The final two videos utilizing first name or nick-name behavior occurred only in title, description or tags. First name or nick-name use was not found in the videos themselves. The sixth video where first-naming occurs is titled [17]. This video has names in the title, description and tags. It is also posted by the username "Tarshh", which is consistent with one of the names in the title and description. The final instance, in [9], also only has the name "paul white" in the title and description.

6.2 Co-operative Decision Making

Co-operative decision making behavior was observed in six out of the ten videos. [2] showed strong co-operative decision making behavior. Throughout the video, scenes emerged and interaction occurred demonstrating this behavior. An example includes the beginning of the video, which depicts three young men discussing the action of chugging beer. The three men agree to drink the beer and begin chugging the beer. During this discussion, other young men off camera can be heard saying "Pete wants to finish it." This discussion leads to the chugging and vomiting behavior found in the video. Similar behavior is observed in [15]. The video features a young man chugging a glass of whiskey, his reaction, and his subsequent behavior. In a scene depicting chugging behavior, the audience can be heard exclaiming "go, go" as he struggles to finish the drink. In a scene in which the young man is being tied up with duct tape, we hear dialogue between the other participants including "keep going" and orders to "lay him on his fucking stomach." In [13], a young man is shown passed out on a couch being marked up with colored markers. The scene opens up with the videographer exclaiming "do it", which is followed by a performer off camera moving to make more marks on the passed out young man's face and back.

During the [17] a constant dialogue between two girls jumping in and out of a bathtub shows co-operative decision making behavior. Another occurrence, in the same video, involves a female off camera instructing the two girls to "go one at a time." In [9], co-operative behavior is taking place between two females who are trying to lift a drunken man out of a bathtub. They work as a team and with the drunk as they lift him out of the bathtub exclaiming "hold on to that, hold on that." [5] features many behaviors that can be interpreted as co-operative decision making. This video features a young man passed out on a couch who is shown carried by two other men into a bathroom and thrown into a bathtub. This entire video depicts co-operative decision making between the performers.

6.3 Profanity

In seven of the ten videos profanity was observed. [2] is set in a dorm room and several times the performers use profanity while observing the situation unfold. In the beginning, one of the actors asks "are you going to finish this, because it's so fucking strong." After chugging the beer, one of the men exclaims "oh fuck" as he belches. Once the young man vomits, several of the actors begin laughing and using profanity as they antagonize the young man for vomiting. In [10], several instances of profanity were tallied. Several times the phrase "oh shit" can be heard from off camera as we observe the young man gagging and running to vomit. After vomiting, the young man also exclaims "oh shit, yeah!" In [1], one person is filmed drinking beer. He begins the video by stating "What up crew… motherfuckin' brew fan." After drinking two beers, the man faces the camera and says "I will be needing this. There's no way my stomach can hold 72 oz of beer in 10 minutes. You gotta let that shit settle."

[15] has several instances of using profanity. When the young man begins to chug a glass of whiskey, someone off camera proclaims "You're fucking sick." In another scene, we see a different man upset saying "You guys are to blame; you guys kept egging him on. All of you kept fucking egging him on". Throughout the video, we also hear people off camera swearing. During [13], we witness a young man being marked on while he is passed out on a futon. While he is being marked on, he awakes enough to kick and slap at the person. During this interaction, he mumbles "fuck off" twice. In [5], profanity is prevalent. Several interactions includes at least one swear word. In the beginning, one young man moves the camera with his hand so it is focused on him and says demonstratively "Fuckin… Hey, Allen's motherfucking becoming 2007 bitches. This is payback for the club." In another scene, we witness two young men picking up another young man who is evidently passed out; the videographer exclaims "Al fucked up in the club, he's about to get fucked." After we see the two young men carrying the passed out individual into a bathroom and throwing him into a bathtub full of water, the cameraman yells "07 nigger". After this, a man off camera states "look at his little pussy ass." The cameraman then

again says "2007 nigger. We ridin' dirty in this bitch." The use of the term "nigger" in this video is considered as risky backstage behavior in our current society. In video [5], all actors are white. To use this term and post it to the Internet is very risky given today's political sensitivity.

The last video utilizing profanity can be seen in [4]. After showing four men chugging beers, we see one of the men start vomiting into a trash can. From off camera, a young man yells "look at these fucking people." Later, while we watch two men vomiting and laughing, a person off camera says "Yo guys, it's too early for this shit."

6.4 Open Sexual Remarks

Only one video contained open sexual remarks. At the end of the [5], while we watch a drunken young man stand shivering in a bathroom after being dropped into a tub of water, we hear a man off-camera say "Hey Allen, I slapped that (incomprehensible) noise the girl made from the time she swallowed man, so chill."

6.5 Elaborate Griping

Two instances were observed involving elaborative griping. In [15], a very irate young man is shown yelling at the entire room. He yells "You guys are to blame. You guys kept egging him on. All of you kept fucking egging him on." The second occurrence of elaborate griping can be seen in [4]. At the end of the video after watching two men vomiting upon chugging beers, a young man off camera yells "What kind of people are we. What kind of people are we." The camera focuses on this young man while people laugh in the background. After a pause, he exclaims "This is ridiculous." In the background a female says "Is it over yet?"

6.6 Smoking

One instance of smoking occurred on camera. In [9], a young woman enters the frame smoking a cigarette. She begins helping another woman remove a drunken man from a bath tub. She turns to the videographer and hands him her cigarette. After handing her cigarette to the videographer, she turns back to the scene and again begins helping the man.

6.7 Rough Informal Dress

In eight out of ten videos we are presented with rough, informal dress. In the videos we see individuals in underwear, swimsuits or some other form of casual attire. Typically we see the person featured in the video in casual wear, although there are instances of other individuals on camera who are in informal dress. In [2], we see men sitting in a dorm room in t-shirts, jeans and backward baseball caps. [6] is a video of a man walking in a field in shorts and a t-shirt. The third occurrence found in [1] shows a man in a hooded sweatshirt and jeans. Half-way through this video the man removes his hooded sweatshirt and is shown wearing a white tank top.

[15] depicts a variety of people wearing baseball caps, blue jeans, and t-shirts. Some are ripped or very worn. At the end of the video we see that the drunken man, who is the focus of the video, wearing no shoes. In the fifth instance [13], we see a young man passed out on a futon wearing blue jeans and a ragged t-shirt that appear to be ripped in several places. The sixth occurrence can be seen in [17]. In this video, we see two young women, inside a house in a bathroom, wearing swimsuits. The seventh instance, in [9], shows two people in their underwear. We see a drunken man in a bathtub wearing a t-shirt and underwear. We also see a young woman who is also wearing only a v-neck shirt and underwear. The eighth instance, in [5], portrays a young man passed out on a couch wearing only a t-shirt and underwear. At the end of the video, we see the same young man after having been dropped in a bath tub full of water wearing wet clothes that are falling down.

6.8 "Sloppy" Sitting and Standing Posture

Because of the nature of the videos retrieved, all ten videos showed instances of "sloppy" sitting or standing postures.

6.9 Use of Dialect or Sub-standard Speech

Again, because of the nature of the videos retrieved, all ten videos showed instances of language that would be deemed dialect or sub-standard in normal interaction behavior.

6.10 Mumbling and Shouting

In nine of the ten videos, there were instances of mumbling and/or shouting. Because videos were examined that contained drinking behavior, this type of verbal behavior is expected. In all nine videos containing this behavior, there were both mumbling and shouting instances.

6.11 Playful Aggressivity

Three of the ten videos observed displayed acts that were deemed playful aggressivity. In [15], we are witness to one act of playful aggressivity. This involves a scene after the drunken young man vomits in which he is duct taped and left on the floor. Everyone is laughing during the scene and the drunken man is taped up and made to put something into his mouth. The second occurrence, in [13], shows a young man who is passed out on a futon being written on by another person. The drunken man kicks and swings at the person marking on him while the other people off camera can be heard laughing. The last video that displays acts of playful aggressivity, [5], presents a young man, who is also passed out, being carried into a bathroom and dropped into a bathtub full of water.

6.12 Inconsiderateness for the Other in Minor but Potentially Symbolic Acts

Depending on interpretation, instances seen in videos could be considered inconsiderateness for the other. However, I chose not to rate the ten videos using this category because of possible inconsistency in interpretation.

6.13 Minor Physical Self-involvement (humming, whistling, chewing, nibbling, belching and flatulence)

Lastly, this category again was seen in all ten of the videos. This is primarily due to the choice of videos to observe. Because all videos contained acts of vomiting and drinking, there were many occurrences of belching throughout the videos critiqued.

7. DISCUSSION

As shown in the previous section, instances of backstage language behavior were prevalent in the videos analyzed. This behavior, although not definitive, provides ample insight into a variety of communication strategies occurring in this subset of YouTube videos. Backstage behavior, typically reserved to members of one's own team outside of the view of the audience, can be seen on these sampled videos

At the beginning of this study, the question was posed: Is there a blurring of the boundaries between front stage and backstage

behavior in videos posted to YouTube? This preliminary work cannot answer this question definitively. There are three possibilities that have emerged utilizing the dramaturgical theory.

This may suggest that the lines are blurring between front stage and backstage behavior. It may also suggest that technology has presented a new communication tool that is not yet governed by traditional communicative patterns. When viewing YouTube videos, we may be gaining access to a communication that was intended for a specific audience. What we gain from the experience may be an insight into backstage behavior of the actor. Goffman suggested that actors define their backstage region based on different situations and that they are always recreating the backstage area. However, in video posts the actor doesn't have the ability to change behavior. Therefore, we may be gaining insight into a particular behavior that was not originally intended for us.

Another possibility could be that the boundaries between front stage and backstage behavior have been moved, allowing previous behaviors defined as backstage to be accepted in the front stage arena. The intent of the videographer may have been to present this behavior in a front stage manner.

Another theory may also exist using the dramaturgical outline. Visitors to YouTube may initially be considered "outsiders." Goffman defines this as region as "neither front nor back with respect to a particular performance… those individuals who are on the outside of an establishment" [7; p. 135]. As outsiders, we are not meant to be the intended audience and therefore could be viewed as having access to the backstage region by simply watching the video. However, this does not seem to be consistent with the norms of Internet behavior. When a person posts a video to YouTube, unless they mark the video as private, they are made aware of the implications. Users of YouTube have only to search for the proper keyword to find any video that may exist.

These possibilities suggest that the question posed in this paper cannot be determined using this small, sample data. Future research is needed to address this issue. This study does indicate that Goffman's dramaturgical theory can be useful when analyzing videos posted on YouTube. Future research should include a larger sample and surveys or interviews collected from the actual videographers to determine intent when posting videos.

"When we blur the boundaries that distinguish private thought from shared experience, when we adjust the lines that separate past, present, and future, or fact from fiction, we expand the confines of what we call reality" [3, p. 55].

8. REFERENCES

[1] Bruz40. *Drinking a 6-pack of Yuengling Lager in 10 minutes*. December 04, 2007. http://www.youtube.com/watch?v=H6v6yZV50sM&feature=related

[2] bwill287. *Post-Chug Puke.* March 04, 2007. http://youtube.com/watch?v=yrTJqfm29Kw&feature=related

[3] Cerulo, K. Reframing sociological concepts for a (virtual?) world. *Sociological Inquiry, 67*,1 (1997), 48-58.

[4] danpat1983. *8am Chug n Puke.* April 09, 2006. http://youtube.com/watch?v=Vnv_lEZyMsQ&feature=related

[5] Drunkproductions . *Drunk Diving*. January 06, 2007. http://www.youtube.com/watch?v=n049WgABHlI&feature=related

[6] d3xt4h. *Guy pukes after funneling a white-beer!* August 10, 2006. http://youtube.com/watch?v=xc6rDN-7Fqs&feature=related

[7] Goffman, E. *Presentation of Self in Everyday Life.* (1973), Woodstock, New York: Overlook Press.

[8] Lenhart, A. Unstable text: An ethnographic look at how bloggers and their audience negotiate self-presentation, authenticity and norm formation. Unpublished master's thesis. Graduate School of Arts and Sciences. Georgetown University, Washington, D.C. (2005). http://lenhart.flashesofpanic.com/Lenhart_thesis.pdf

[9] medionmike87. *Drunk Paul White from Midd.* May 24, 2007. http://www.youtube.com/watch?v=BHRVpG-UqXs&feature=related

[10] onlinepredator3. *Vince Vomit Puke Beer Chug.* April 11, 2006. http://youtube.com/watch?v=Rr_0Ncqw6pg&feature=related

[11] Ross, D.A. Backstage with the knowledge boys and girls: Goffman and distributed agency in an organic online community. *Organization Studies, 28*,3 (2007), 307-325.

[12] Scheff, T.J. *Goffman Unbound! A New Paradigm for Social Science.* (2006), London: Paradigm Publishers.

[13] scottyboy278. *You Thought Red Hair was Bad, This is Worse.* October 01, 2007. http://www.youtube.com/watch?v=D_AhkhBiDrw&feature=related

[14] Shapiro, E.J. & Shapiro, L.D. *Presentation of Self in Virtual Life Web Page Personas in Cyberspace.* (1997) http://www.cs.pdx.edu/~len/webpage.doc

[15] SteveSS666. *Parsons drinking CC.* September 26, 2007. http://www.youtube.com/watch?v=-xHbZ6fZsjY&feature=related

[16] Tardy, R.W. "But I am a good mom": The social construction of motherhood through healt-care conversations. *Journal of Contemporary Ethnography*, 29 (2000), 433-473.

[17] Tarsh. *Tarsh N Rach drunkkkk!* October 01, 2007. http://www.youtube.com/watch?v=EpJmxxDsQEM&feature=related

[18] YouTube. *YouTube Advertising Opportunities*. (2007) http://www.youtube.com/advertise

Sickness and Health: Homophily in Online Health Forums

Brant Chee
University of Illinois at Urbana-Champaign
Graduate School of Library and Information Science
chee@uiuc.edu

ABSTRACT
This work explores the link between health and social relations by creating an automated metric of similarity of positive or negative affect (sentiment) between peers in online health forums. We analyze textual communication between peers and demonstrate that those who communicate often have similar average sentiment scores. Sentiment is the author's immediate affective state, their positive or negative orientation. We hypothesize that average sentiment over time indicates overall happiness or sadness as similar analysis has been utilized to identify depression and depression at risk college students [19]. These results follow the analysis of Framingham study data demonstrating that happy people tend to associate with one another and that happiness spreads within social networks [4].

Categories and Subject Descriptors
J.4 [Social and Behavior Sciences]: Sociology

General Terms
Measurement

Keywords
Sentiment Analysis, Social Networks, Tie Strength Indicator, Homophily

1. INTRODUCTION
Happiness and depression are important factors for chronic illness patients. The Centers for Disease Control (CDC) currently uses a 14 item measure based upon patient response to simple questions including topics on stress, depression and problems with emotions [13]. The CDC believes that health-related quality of life (HRQOL) is important in the measurement of effects of chronic illness on patient's lives. HRQOL is important in tracking patient's perceived physical and mental health over time and tracking the effects of multiple diseases and disabilities within patient populations [13]. With further work this average sentiment metric could provide an automatically generated measure to augment existing HRQOL measurement tools.

Homophily is the principle that similar people – in many regards to socio-demographic, behavioral, and intrapersonal characteristics including race and ethnicity, interact more than those who are dissimilar [14]. The goal of this work is to derive a quantifiable metric for automatically measuring the similarity of people with respect to their attitudes, abilities, beliefs and aspirations. Specifically, we are looking at features pertaining to value homophily. This includes the wide variety of internal states presumed to shape our orientation toward future behavior. Examples of values might include higher education attainment, social characteristics that can be correlated with political similarity, etc. People tend to assume that their friends are like them [14]. People's social network, their relationships and interactions with other people, is formed by whom they choose to interact with.

Within the social networking paradigm people represent the nodes in a social network diagram or sociogram and their interactions represent the edges connecting the nodes. Patterns of interaction are used to demonstrate an effect, such as, people whose friends become obese tend to also become obese [2]. In many cases the network and effect are hard to construct due to the difficulty in ascertaining necessary information about social interaction and the hypothesized social effect. Special datasets such as the Framingham study are often used. This dataset includes manually collected data over a period of 20 years, preventing the analysis from being performed on many people [2,3,4]. We would like to perform similar sociological analysis on new segments of the population automatically.

In this work we define a similarity metric for a person's cognitive model and overall sentiment utilizing the words they use when constructing messages to other people. This sentiment metric is used as a measure of a type of value homophily within the Yahoo Health forums. We demonstrate that people's affective or emotional state, specifically their positive or negative orientation, is likely to be similar to others' they choose to associate with.

For our analysis, we use messages within online Yahoo Health forums, constructing a social network through people's message exchanges. Email is used as a proxy for measuring relationship strength, which others have done previously [20]. We demonstrate that people's affective emotional state is similar to others they communicate frequently with.

2. RELATED WORK
The words people use in conversation correlate to physical and mental health [19]. Research in content analysis introduced in the 1960's detected a person's affective or immediate feeling state based solely on variations in the content of verbal communications [8]. The same language processing technique was used in the late 1970's to differentiate between people with schizophrenia and those without [18]. Following related work focused on written text, finding variations in language usage between depressed, depression-vulnerable and non-depressed students [19]. Much work has been done on the automatic detection and analysis of sentiment [15]. One way to think of sentiment is an author's attitude; the positive or negative polarity apparent through the author's writing.

Similar work by Fowler and Christakis addressed the spread of happiness within a large social network. The same dataset has been used to demonstrate smoking cessation through one's social network [3]. Our goal is to automatically compute results for large populations of users and datasets. Little work has been done on the construction of social networks within the context of online

health forums. Furthermore, this work specifically addresses the idea of automatically constructing and analyzing these networks utilizing the concept of homophily.

Adamic and Adar look at websites, specifically MIT and Stanford student home pages to programmatically create a social network, and from it predict friendships [1]. The social network is automatically constructed from web pages (not messages). Verification of friendship is accomplished manually. While homophily is not addressed, the idea that, the "more similar a person are, the more likely they are to be a friend" is. The measure of friendship is determined by text, links, and mailing lists.

Work by Golder et al used Facebook messages to find temporal rhythms consistent across university campuses and seasons [7]. They demonstrate that students at the same university have similar messaging habits. No content analysis was performed on the messages and the resulting data was not used to construct a social network.

3. EXPERIMENT

3.1 People and Data

The social forums we explored consist of 27,290 public Yahoo Health groups. 12,519,807 messages exist within these groups. These groups range from illness based support groups focusing on Multiple Sclerosis to groups focusing on herbal home remedies. For this study, we looked at the 10 largest message groups by file size from this Yahoo corpus.

The messages in these forums consist of informal, often emotional text dealing with feelings of hopelessness, depression or bereavement, for example, "My doctor told me that it works for both depression and as an antianxiety drug… I was in such adepressed stat that I had to go for counselling. [taken verbatim]" Recent studies have shown that the expression of emotional experiences either verbally or in a written context leads to improved physical and psychological health [16]. These texts can also provide emotional insights about the author's mental state at that point in time [19].

Words are not the only elements of analysis that provide necessary emotional insights. People augment computer mediated communication to mimic face to face interactions through the use of nonverbal elements [21]. Emoticons are nonverbal expressions and are often textual representations of writer's facial expressions [5]. For example :) or :-) would correspond to a smile indicating happiness. These cues indicate to the reader the author's intensions which can be hard to determine in informal written communication.

3.2 Experimental Design

We seek to quantitatively determine if a person's affective or emotional state, specifically their positive or negative orientation, is likely to be similar to others' they choose to associate with.

A social network from the messages within the Yahoo group was constructed for this experiment. From this network we extract pairs of people who have numerous interactions and calculate the difference in sentiment scores between the nodes. The same number of pairs is selected at random from the network and the difference in sentiment scores between these pairs of random nodes is calculated. Statistical tests on the two groups (random pairs and interacting pairs) of differential sentiment scores are performed to determine if there is a significant difference between them.

The Yahoo groups were first parsed such that non-text including images or attachments and replies were removed. From the messages, we create a social network. Within the network, nodes are email addresses, which serve as a unique identifier for a person. An edge is formed between nodes *A* and *B* if *B* responds to a message posted by *A* or vice versa. We arbitrarily decided that an interaction of ten messages between two people implies a strong tie. Figure 1 demonstrates an example of a social network. The strong ties are solid black and consist of ten or more interactions. The dotted lines represent weak ties.

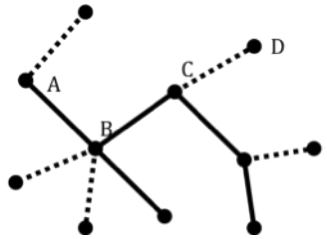


Figure 1: Example social network, strongly connected nodes are represented by solid black lines e.g. nodes A and B. Weakly connected ones e.g. nodes C and D are represented by dotted lines.

The correlation between strong ties, in this case numerous communications between *A* and *B* and sentiment is interesting. If they co-vary together, this suggests homophily. While replies help us understand the context of a message, a message's emotional circumstance should not be based on what other people write, only on the authors' text. Therefore, we remove replies from the content of messages.

Portions of the lexicon in the Linguistic Inquiry and Word Count (LIWC) were used [17]. Specifically, the words corresponding to the following categories: positive emotion, negative emotion, anxiety, anger and sadness. We have augmented the LIWC lexicon to include a wide range of emoticons such as :) :(:P ^__^ LOL ROFL. The resulting messages were matched against the LIWC lexicon categories and emoticons. Counts containing number of positive emotion words, and negative ones, and total number of words were recorded.

A score for each message was calculated. In the score calculation shown below, the set of positive terms denoted as *p* consist of the positive emotion words including positive emoticons. The set of negative terms denoted *n* consists of the negative emotion, anxiety, anger, and sadness terms from LIWC as well as negative emoticons such as :(.

Each message *m* consists of the set of *t* white space delimited tokens such that $m=\{t\}$.

$$score = \frac{\sum_i (\alpha \mid i \in p) - (\alpha \mid i \in n)}{|m|}$$

For each token *i* in a message *m* if *i* is in the set of positive LIWC lexicon α is added to the score and if *i* is in the negative LIWC lexicon α is deducted from the aggregate score. The aggregate score is then divided by the number of tokens. Here α is a constant in this case $\alpha = 1$; thus a message consisting primarily of positive lexicon will have a score > 0, whereas a purely

informational message will have a score = 0, and a message that is predominantly negative will have a score < 0. For each person within the network, an aggregate score is calculated by taking the average score of each of the messages they have written.

The absolute value of the difference in average sentiment between people with strong ties was calculated. The difference in average sentiment provides a distance metric between the two people. We call this metric sentiment distance. We hypothesize that pairs of nodes for example (*A,B*) and (*B,C*) in Figure 1 will have a smaller sentiment distance than randomly chosen nodes. Intuitively, this means that because people, (nodes) *A* and *B* chose to communicate more, they are more likely to have similar positive or negative alignment.

However, in calculating the sentiment for each person who has met the communication threshold or "strong tie", we do not use the messages where the people communicated. We did this so that the language and topic of the threaded messages did not bias the average score.

For every pair of strongly connected nodes, for example (*A*,*B*) and (*B*,*C*) in Figure 1, we pick two other nodes at random which are not strongly connected, for example (*C*,*D*) or (*A*,*D*). We calculate the sentiment distance between the pairs of nodes. A sentiment distance can be calculated between every pair of nodes within the network. The nodes do not need to be directly connected in the network in order to calculate the sentiment distance since it is based on the messages of the people represented by those nodes.

This process of choosing pairs of nodes and calculating the sentiment distance creates two distributions named strong and weak. The strong distribution is composed of the pair wise sentiment distances between people who have had more than ten communications. The weak are the same number of pairs of randomly selected nodes. [11] show that for a student's t-test sample populations should be approximately equal. Considering all nodes *n*, in a graph, and *s* is the set of strongly connected pairs, the number of random pairs, $nr = (n\text{-}1)! - |s|$ where $(n\text{-}1)! >> |s|$ in the general case, so sampling was used.

4. RESULTS

The mean message size for each group is 160,984 messages. The table below lists the statistics for the different groups. The number of pairs of nodes with numerous interactions (10 or more) is shown in the column second from the left. The average distance between those nodes with numerous interactions is in the Strong column followed by the average distance between random pairs of nodes in the Weak column. The p-value is the result of a T-test comparing the distribution of the Strong column values to the Weak column ones. The last column shows the ratio of the average difference between the two populations to demonstrate the quantifiable difference between the two. For each group the mean difference between the sentiment of strong pairs and weak ones appear to be statistically significant with p-values much lower than .05. The people who have numerous interactions are much closer in average sentiment than pairs chosen at random, thus indicating their mental model and happiness levels are similar.

The average ratio of the mean differences between the population of random people versus those with numerous interactions is 2.152, not only is the average difference between nodes with numerous interactions significant, it is less than half of that between those with few interactions.

Of particular interest is that the network was automatically generated from people's behavior within groups. The metrics we use to define similarity are automatically derived and demonstrate that such automatic metrics are still able to detect these similarities with great reliability.

Table 1: Table displaying the means and p-values for T-tests comparing strongly connected people and weakly connected ones within ten Yahoo! Health forums.

Group	Pairs	Strong	Weak	p-value	Weak/ Strong
1	888	0.0082	0.0144	2.20E-16	1.756
2	505	0.0104	0.0324	2.20E-16	3.115
3	463	0.01	0.0243	2.20E-16	2.430
4	398	0.0171	0.0281	1.37E-11	1.643
5	380	0.0058	0.0129	2.20E-16	2.224
6	341	0.0173	0.0207	3.69E-03	1.197
7	306	0.0096	0.0169	2.28E-09	1.760
8	262	0.0112	0.031	2.20E-16	2.768
9	237	0.0096	0.0199	2.20E-16	2.073
10	233	0.0074	0.0189	2.96E-12	2.554

4.1 Limitations

These finding validate the idea that people who interact frequently have similar average sentiment and therefore similar mental models, however these results do not indicate causality. Our current work only explored the differences between an arbitrarily defined strong tie and a weak one. It is not known if the current metric is a continuous one, which would prove more useful as Gilbert's work suggests [6]. We do not currently have any information on how sentiment between people is affected by the strengthening or dissolution of ties.

4.2 Implications

4.2.1 Practical Implications

People with strong ties have a small difference in their average sentiment scores. A potential implication is use of average sentiment difference as a feature in the calculation of tie strength. Gilbert and Karahalios used words in inbox messages and Facebook wall posts to quantify tie strength, however they did not look at average sentiment of posts and message [6].

Average sentiment analysis is more computationally expensive than finding people with numerous communications. Utilizing our findings, it is possible to create groups of people utilizing this cheaper distance metric and verify their average emotional distance to create groups of people with similar value homophily. Implications of similar value and emotional/psychological-based groups include targeted advertising, identification of depression at-risk populations. Previous work by Rude et al. shows that automatic detection of people with depression is possible [19].

4.2.2 Theoretical Implications

Hancock et al. demonstrate that emotional contagion, the mood of one person can change the mood of others interacting with that person in text-based communication [10]. Similar results were demonstrated for groups of people in a Social Network and shifts in happiness of people within that group [4].

People who are optimistic tend to be healthier and live longer than those who are pessimistic and cynical. A long term study started in 1921, of 1,500 pre-adolescent boys demonstrated that expecting the worst was linked to a 25-percent higher risk of dying before age 65 [12]. Over 1300 people in a 10 year Harvard study showed cardio-protective effects of optimism; the risk of coronary death or disease, Angina, or non-fatal Myocardial infarction was reduced by half [9].

This work may contribute to the development of a quality of life metric utilizing average positive or negative orientation. Further, since one's orientation changes depending on whom one interacts with, we conjecture it is possible to change a person's orientation by changing who they interact with. Possible implementations of this include re-ordering of people's information, messages in forums to rank negative people's posts higher, or to suggest friends who are positive for negative oriented people of whom have weak ties connecting them.

5. ACKNOWLEDGMENTS

This work was an independent study oversaw by Caroline Haythornthwaite and was partially supported by a fellowship from the Center for Integration of Medicine and Innovated Technology. We thank reviewers for helpful comments.

6. REFERENCES

[1] Adamic, L.A., and Adar, E. 2001. Friends and neighbors on the web. Social Networks, 35 (2001), 211-230.

[2] Christakis N.A. and Fowler J.H. 2007. The spread of obesity in a large social network over 32 years. N Engl J Med, 357 (2007), 370-379.

[3] Christakis N.A. and Fowler J.H. 2008. The collective dynamics of smoking in a large social network. N Engl J Med, 358 (2008), 2249-2258.

[4] Fowler JH, Christakis NA. 2008. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study. BMJ, (2008) 337a2338.

[5] Gajadhar J. and Green J. 2005. An analysis of nonverbal communication in an online chat group. EDUCAUSE Quarterly 24, 4(2005), 63-64.

[6] Gilbert, E. and Karahalios, K. 2009. Predicting tie strength with social media. 2008. In Proc. CHI 2009, ACM Press (2009), 211-220.

[7] Golder, S., Wilkinson, D., and Huberman, B. 2007. Rhythms of social interaction: Messaging within a massive online network. In Proc. CT2007, Springer (2007), 41-66.

[8] Gottschalk L.A. and Gleser G.C. 1969. The Measurement of Psychological States through the Content Analysis of Verbal Behavior. University of California Press, Berkeley, CA.

[9] Kubzansky, L.D., Sparrow, D., Vokonas, P., Kawachi, I. 2002. Is the glass half empty or half full? A prospective study of optimism and coronary heart disease in the normative aging study. Psychosom Med, 63 (2002), 910-916.

[10] Hancock, J.T., Gee, K., Cicaccio, K. and Lin, J.M.H. 2008. I'm sad you're sad: Emotional contagion in CMC. In Proc. CHI 2008, ACM Press (2008), 295-298.

[11] Markowski, C. A. and Markowski, E. P. 1990. Conditions for the effectiveness of a preliminary test of variance. The American Statistician, 44, 4 (1990), 322–326.

[12] Maruta, T., Colligan, R.C., Malinchoc, M., Offord, K.P. 2000. Optimists vs pessimists: Survival rate among medical patients over a 30-year period. Mayo Clin Proc.,75, 2 (2000), 140-143.

[13] Moriarty, D. B., Zack, M. M., and Kobau, R. 2003. The centers for disease control and prevention's healthy days measures - population tracking of perceived physical and mental health over time. Health and Quality of Life Outcomes (2003), 1-37. DOI=10.1186/1477-7525-1-37

[14] McPherson M., Smith-Lovin L., and Cook J.M. 2001. Birds of a feather: Homophily in social networks. Annu. Rev. Sociol., 27 (2001), 415-444.

[15] Pang B. and Lee L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2, 1–2 (2008), 1-135.

[16] Pennebaker J.W. and Campbell R.S. 2000. The effects of writing about traumatic experience. Clinical Quarterly, 9 (2000), 17-21.

[17] Pennebaker J.W., Francis M.E., and Booth R.J. 2007. Linguistic inquiry and word count: LIWC 2007. Lawrence Erlbaum Assoc, New Jersey, 2007.

[18] Rosenberg S.D. and Tucker G.J. 1979. Verbal behavior and Schizophrenia. Arch Gen Psychiatry, 36 (1979), 1331-1337.

[19] Rude S.S., Gortner E.-M., and Pennebaker J.W. 2004. Language use of depressed and depression-vulnerable college students. Cognition and Emotion, 18, 8 (2004), 1121-1133.

[20] Tyler, J., Wilkinson, D., and Huberman, B. 2003. Email as spectroscopy: Automated discovery of community structure within organizations. In Communities and Technologies, M. Huysman, E. Wenger, and V. Wulf, Eds. Kluwer, Deventer, the Netherlands, 81-96, 2003.

[21] Walther J.B. and D'Addario K.P. 2001. The impacts of emoticons on message interpretation in computer-mediated communication. Social Science Computer Review, 19, 3 (2001),323–345.

Social Support in Online Healthcare Social Networking

Katherine Chuang
iSchool at Drexel
3141 Chestnut St.
Philadelphia, PA 19104
katherine.chuang@ischool.drexel.edu

Christopher C. Yang
iSchool at Drexel
3141 Chestnut St.
Philadelphia, PA 19104
chris.yang@drexel.edu

Abstract

Support groups play an important role in helping individuals cope with their health conditions by providing members with a network of information and social support. It is not well understood what types of support are exchanged in an online, unmoderated, peer-to-peer communication of patients and caregivers. We analyzed the informational and emotional support offered and requested among an alcoholism discussion group using qualitative content analysis. Members offered a wealth of information in the form of facts, advice, and personal stories. Emotional support was also prevalent among threads where it was not explicitly requested. Our results highlight how peers are supportive to one another by providing resources and encouragement. These findings bring insight to new support we could provide to patients for sharing information, such as health professionals developing intervention programs. This work is a piece of an ongoing project that identifies the communication patterns of patients in online support groups.

1. INTRODUCTION

Online social networks allow users to connect with each other by overcoming geographic and time boundaries. Patients or their caregivers often turn to the Internet to seek support (i.e. advice and words of wisdom, sharing experiences), especially those who lack local resources. ñE-patientsò increasingly use social networking platforms to teach each other about conditions and treatments. This will impact the traditional patient-doctor relationship, and could even create the basis for a more market-driven system where customers are able to make informed choices [18]. Knowing what e-patients are seeking from these websites can provide insight to health professionals designing intervention programs, or inform design of these communities.

An online community such as a social networking site is a place where registered users create profiles about themselves, upload photos, keep in touch with friends and make new friends with common interests [4]. Medhelp is one such free online health community that connects people with medical experts and others who have similar experiences. Each day, members visit MedHelp to receive the support they need from other patients like them, and to share their knowledge with others in need. The website features news, clinical trials, personalized tools and chat forums. It was founded in 1994 as a resource to help patients cope with their health conditions by connecting users with information resources. It now hosts over a million users and hundreds of patient communities that can communicate through forums, profiles or journals.

Online support communities sometimes supplement regular offline group meetings (e.g. for people recovering from alcohol abuse) because it can provide additional emotional support for participants in offline recovery programs [19]. [24] studied the impact of participation in online and offline support groups on the perceived stress level and coping strategies of participants. He found that the more time people spend in an online group, the larger their online social network and the higher the satisfaction with the received support. [12][13] found that the Internet was a potentially powerful tool in motivating and assisting problematic drinkers due to its anonymity and accessibility, where participants may not feel the same stigma associated with participation in face to face intervention [15]. Peer communication can play a role in facilitating new health habits, such as quitting smoking [3].

Psychology and communication research on social support shows two forms: resources intended to assist stressed individual to solve or eliminate problems causing distress, and emotional understanding to comfort support seekers [9][14]. Other types of positive interactions common among online support groups include introductions, expressions of gratitude or congratulations [7][10][11][21]. The first type, action-facilitating support, solves or eliminates problems facing support seekers by providing information or tangible support. Information resources include: offering advice, information referral, insight from personal experiences, or opinions **Error! Reference source not found.**[7][8][10][14][15]. Instrumental support includes offering financial assistance, services to relieve stress, active participation, or willingness to help [1][7][8][10][14][15]. Secondly, nurturant support comforts stressed individuals without direct efforts to solve problems causing the stress by making the recipient feel cared. Using verbal and nonverbal communication, giving compliments and recognizing achievement, and to help him/her have a sense of belonging among people with similar concerns [1][7][8][10][14][15]. Several studies have investigated social support exchanges in electronic environments for different patient communities such as breast cancer [8], disabilities [5], HIV/Aids [10], eating disorders [15], psychosis [7], and depression [21]. In this work, we look at the same range of informational, emotional and instrumental support in an online alcoholism community.

Moderators can be beneficial members of an online community [11]. In their study, they analyzed activity over a nine-month period on AlcoholHelpCenter.net, a moderated alcoholism community and found common themes in messages include (ordered from most to least): introductions, greetings, general supportive statements, suggested strategies, success stories, and discussion of difficulties. There were few extremely active users, but of the members who posted at least one message, over half were female. Their analysis examined types of messages rather than the prevalence support offered and sought. [6] conducted a qualitative content analysis of an online support group for smokers and concluded that the support group was mainly used as a source of support and encouragement during the initial phases of quitting cigarettes. Practical information and quitting tips were

less common, unlike in the alcoholism community where users were likely to describe why they decided to change habits [11].

The objective of this preliminary study is the first part of an ongoing project to use an unmoderated alcoholic community to understand the prevalence and characteristics of different support types (informational, instrumental, nurturant) that peers exchange in an online forum. This understanding can inform development of support for individuals to share experiences for overcoming alcoholism. Knowing characteristics of social support offered and sought in these virtual groups will help us to better understand the range of information needs (i.e. advice, sympathy) in a supportive environment. The contribution from such a study would provide greater insight into how alcoholics and their caretakers participate in online forums, or other means of communication. In the next phase of this ongoing project, we shall investigate the social supports in other channels such as journal and notes. We shall identify if the types of social support change when users of the same community interact through a different communication format. Although we select the alcoholism community in this study, we shall investigate if the types of social support change across different disease community in the future.

2. APPROACH

We chose the discussion board about alcoholism from MedHelp (www.medhelp.org) for this study. This study adopts a descriptive view to identify the different types of social support based on definitions given from relevant literature. A web crawler downloaded all the forum posts through 9th September 2009. There were 737 threads overall written by 454 individuals. A pilot study of threaded 493 forum messages was selected from a three-month period (9th June 2009 to 9th September 2009). This sample size of 493 messages is large enough to do a meaningful analysis. The sample is composed of 81 thread posts and 412 comments to the posts. Threads were initiated by 68 users. 97 users participated in posting comments. Some users had initiated threads and posted comments. A total of 128 participants had involved in the forum. The average number of comments per post is 5.08. The average number of participating users per thread is 4. The maximum number of participating users per thread is 12 and the median number of participating users per thread is 3.

The goal of this research was to find examples of social support as they occur in online communication. Qualitative content analysis was chosen since it deals with the interpretation of the codes. We followed an inductive approach to determine the different types of social support under categories: Informational, Nurturant, and Instrumental. In addition, we looked at interactions across several dimensions (offered vs. received support, post vs. comments). Offered support refers to the support that is given or expressed in the message, i.e. ñ*you should see your doctor to address your condition*ò. Requested support refers to the type of support sought, such as ñ*is there a medicine to take to stop the craving for alcoholic drink*?ò The threads are arranged by the website into the first post of a thread and the comments. The following figures show example of post and corresponding comments.

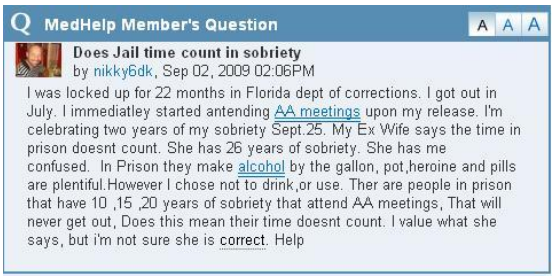


Figure 1 Example of MedHelp forum post

Figure 2 Example of MedHelp forum comments

Development of code scheme

Categories for coding were developed by reviewing existing samples and descriptions in related the research studies, organized in the categories defined by [14]. These categories are appropriate for this study and they were also used in other studies of online support groups [5][7][10][15].

Analysis

We calculated the frequencies of each code category to analyze the overall levels of activity of each post (i.e., how many codes per message?) as well as the support categories. The most support, *Informational Support,* was further analyzed in sub types. We ignore typos and grammatical errors in messages.

Coding Scheme

There are 10 codes sorted into three categories. The categories are not exhaustive, in that some of the behaviors noted by literature did not match this data (e.g., instrumental support). Only those instances that fall into one of the 10 social support categories were tallied.

Informational Support: posts providing information on treatment, coping, etc. [1][7][8][10][14][15] Subcategories include:
Advice: Offers ideas and suggests actions, Provides detailed information, facts, or news about the situation or skills needed to deal with situation [14].
Example: *"Campral works better...ask u r doc about it!"*
Suggestions for course of action tell someone in order to improve the situation, such as speaking with a doctor or joining a face-to-face group.
Referral: Refers the recipient to some other source of help [14].

Example: *"Im gonna send you a link that might help."*
This is directing someone to a resource such as book or website that contains related health information, rather than telling someone to take a course of action.

Fact: reassesses the situation and presents facts [14].
Example: *"Drinking too much alcohol daily can be a high risk to your health, you might fall into alcohol addiction."*
This includes telling someone information that can be supported by evidence or unbiased observations.

Personal experiences: stories about personõs experiences [14]
Example: *"I have been going though something like that with an addict using in our bathroom and…."*
Personal experiences are different from facts in that it contains a personal nature, such as descriptions of specific events not medically related.

Feedback/Opinions: a view or judgment formed about something, not necessarily based on fact or knowledge [8]).
Example: ñ*From what have you posted, it seems that you are in the stage where you have been looking to drink everyday and it is a clear sign of alcohol addiction."*
This is different from advice because it does not provide a suggestion for action, but rather thoughts on a situation.

Nurturant Support: posts providing expressions of caring or concern. [1][7][8][10][14][15] Subcategories include:
Esteem: positive comments to praise support seekers abilities or to alleviate feelings of guilt [9][14].
Example: *"Congratulations on your sobriety!"*

Network: messages to broaden support seekers social network so they donõt feel alone [14].
Example: *"Just reach out and I will be there ok?"*

Emotional: providing understanding of situation, express sorrow, provide with hope and confidence [8][14].
Example: *"You're going through a rough time...."* or *"Hang in there hon"*

Instrumental Support: provision of material or financial aid, or services. [1][7][10][14][15]. There are no examples available from the dataset. An example is offering to drive someone to Alcoholics Anonymous meeting.

3. RESULTS

Among the 81 threads in the sample, 56 postings were created by users seeking support for themselves; 14 by caretakers (of friends (3), family members (7), spouse (2), or by significant other (2)); and 11 unknown. There were 412 comments to these posts, which totals to 493 messages. Support is coded for both those offered and sought, and by first post of each thread and the comments.

Table 1. Summary of Offered Support

	Posts	Comments	Combined
Information	82.7 (67)	85.2 (351)	84.7(418)
Nurturant	16.0 (13)	66.9 (276)	58.6 (289)
Instrument	0.0 (0)	0.0 (0)	0.0 (0)
Sample Size	81	412	493

Of the offered support, 67 posts out of 81, or 82.7%, were providing information. For instance, users introduced themselves by describing how much they drink or stories of how alcohol is disruptive to their lives. Users were less likely to start threads offering nurturant support (16%) such as spending time together. Some messages had instances of both information and emotional support. 85.2% of the comments offered information such as updates on a situation or answering questions in the posts. 66.9% of comments offer emotional support such as sympathy.

Table 2. Different types of support offered

# offered	Posts	Comments	Combined
0	8	4	12
1	37	102	139
2	28	129	157
3	7	106	113
4+	1	71	72
Sample	81	412	493

8 posts and 4 comments did not offer any type of support. Most messages had one or two codes, while many messages multiple types of support, especially in the comments.

Table 3. Summary of Requested Support

	Posts	Comments	Combined
Information	72.8 (59)	15.5 (64)	24.9 (123)
Nurturant	44.4 (36)	6.3 (26)	15.0 (62)
Instrument	0.0 (0)	0.0 (0)	0.0 (0)
Sample Size	81	412	493

59 posts out of 81 (72.8%) sought information such as recommended drugs for treatment. 44.4% sought emotional support from the forum. Conversely, in the comments there were fewer instances of requested support. 15.5% sought informational support such as clarification and a minimal 6.3% were looking for emotional support or validation.

Table 4. Different types of support requested

# requested	Posts	Comments	Combined
0	22	349	371
1	41	58	99
2	16	5	21
3	2	0	2
Sample	81	412	493

In the data, most of the messages requested minimal support. 22 posts and 349 comments did not request support. Few messages requested more than one type of support and that is more likely to occur in posts than comments.

Each message on average contains 2.57 support codes. This number includes both offered and requested support. The maximum number of codes per message is 10, except among 1st post of each thread, which had maximum of 6 codes. Some messages only offer support (i.e., ñ*Have you tried Naltrexone? It is suposed to help with the cravings there are other meds that can help with it too. If all else fails, make a picture of tea and pop some popcorn and hang out with him with your "drink*ò), or only request support (i.e., ñ*Hi, is there a medicine to take to stop the craving for alcoholic drink?*ò). *Factual Information* was the most frequent category of the code scheme, present in 31.2% of the messages. An example of this would be a statement describing consequences of being an alcoholic, ñ*Suboxone is not for alcoholics dear! it is primarily used for heroin and opiate addicts who have used heavily for a substantial length of time." or "I'm 52, been drinking heavily since I was 27".* This shows how important it is for members of this online community to exchange information related to their situation (e.g., how long they have been drinking or sober, how often they drink, use of medicine, etc). The information support was observed in over half the messages (59.0%), where *Advice* (12.3%), Personal *Experiences* (8.0%), and *Fact* (31.2%) occur most frequently. The categories *esteem* (5.6%) and *network presence* (3.3%) occurs least frequently.

Information support was the most frequently noted category in this data set, so we further analyzed the types of information exchanged in this forum. Within the information support category, the table below shows frequencies of offered and requested information support in messages. Other researchers have noticed that people often ask for opinions in addition to wanting factual information so we also noted the occurrence of opinions in this sample [2].

Table 5. Informational Support

	Posts		Comments	
	Offered	Requested	Offered	Requested
Advice	0% (0)	27.1% (22)	32.5% (134)	1.9% (8)
Referral	0% (0)	4.9% (4)	4.9% (20)	0% (0)
Fact	74.0% (60)	48.1% (39)	64.8% (267)	12.3% (51)
Personal	33.3% (27)	1.2% (1)	18.7% (77)	0.5% (2)
Opinion	0% (0)	16.0% (13)	13.6% (56)	1.7% (7)

4. DISCUSSION

This paper describes the results from a preliminary study of an online alcoholism community. Content analysis coding scheme developed from literature organized into Cutrona & Suhrôs category system [14]show that users of this community are similar to other communities who offer and seek general informational and emotional support [7][8][10][15][21]. This seems logical since people participating in support groups want to improve their situation[23]. Our results support previous findings by similar studies using same support categories developed by [14] who found little or no instrumental support [10][15]. Other studies looking at online support communities using their own coding schemes also did not find instrumental support [8][11][21]. We observed no instrumental support in this study. This may be an attribute of virtual environments, where users do not know each other and may be located at diverse geographic distance. When two people know each other well, such as in the case of a married couple, instrumental support is more likely to occur [14]. Some messages may contain very few codes when user goes off topic; this is consistent with previous findings [15].

Our findings support previous studies showing greater informational support than the other types [7][10][15]. Providing information may be a strategy of support seeking by self-disclosure [15]. Themes in messages emerged about sharing personal experiences such as success stories, discussion of difficulties and exchanging advice, and general supportive statements. One main difference between our study and that of [11] is that MedHelp is an unmoderated community while AlcoholHelpCenter.net uses moderators, whose role is to greet new members, give general encouragement or specific suggestions. In the MedHelp alcoholism community, several of the most active members stepped up to this role. [8] found that patients exchange expertise in the form of describing types of problems, what to expect in different situations, solving the problems, recommendations, personal stories, suggested approaches, or referrals to information resources. Our findings suggest that ex-alcoholics offering personal stories and suggestions have valuable insight to helping alcoholics. Empathic interactions in online communities include self-disclosure, community building, medical facts, technical issues, or off topic stories [21]. The presence of emotional support is influenced by the topic being discussed, presence of women, and presence of moderators [22]. The presence of greater informational than emotional support found in this online support group may be explained by the nature of health-related groups, which are more likely to be fact-based [16].

5. CONCLUSION

The objective of this study was to identify the various types of support present in an online alcoholism community. Peers of online alcoholism communities are likely to exchange information and emotional support, even without moderators for discussion. Information support centered on facts about alcoholism, such as treatment or medicines. The advice peers recommended and personal experience shared reflect patient expertise they have gained through their own alcoholic experience and offer a source for others to learn from similar situations. Overall, it appears that participating in an online discussion board is therapeutic and constructive for individuals with alcohol addiction.

These findings bring insight to the range of support that could be provided to individuals for disseminating information in intervention programs. For example, health professionals can target users who are likely to share information and train them to give the right information. Doing this kind of work can provide many opportunities for future research, for instance, threads could be grouped for improved online discussion forum browsing experience.

Learning about the nature of social support provides a tremendous opportunity to characterize the kinds of assistance that peers can provide. Studies such as ours provide a new view of patient communication that health professionals could utilize to reach out to the public. Future research could explore the motivations

behind postings and the actual effectiveness of these social supportive messages.

The MedHelp community includes other features for patients to interact with each other, such as journals and notes. Journals allow users to record their thoughts and feelings, just like a personal diary. Sharing is can be set to viewable by anyone, limited friends or private. Notes are a way for users to keep in touch with each other, such as congratulating or saying hello. They can be viewed by everyone, friends, or no one based on settings. In future work we will be analyzing social support types on these additional communication channels.

6. REFERENCES

[1] Adamic, L. A., Zhang, J., Bakshy, E. and Ackerman, M. S. Knowledge sharing and yahoo answers: everyone knows something. *Proc. of the 17th international conference on World Wide Web* (Beijing, China, 2008). ACM

[2] Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G. Finding high-quality content in social media. *Proc. of the international conference on Web search and web data mining* (Palo Alto, California, USA, 2008). ACM

[3] Ancker, J. S., Carpenter, K. M., Greene, P., Hoffman, R., Kukafka, R., Marlow, L. A. V., Prigerson, H. G. and Quillin, J. M. Peer-to-Peer Communication, Cancer Prevention, and the Internet. *Journal of Health Communication: International Perspectives*, 14, 1 supp 1 (2009), 38 - 46.

[4] boyd, d. m., & Ellison, N. B. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13, 1 (2007), article 11. http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

[5] Braithwaite, D. O., V. R. Waldron, et al. Communication of Social Support in Computer-Mediated Groups for People With Disabilities. *Health Communication* 11, 2 (1999), 123 - 151.

[6] Burri M., Baujar, V., Etter, J.F. A qualitative analysis of an Internet discussion forum for recent ex-smokers. Nicotine Tobacco Research. 8 (2006), S13-9.

[7] Chang, H.-J. Online Supportive Interactions: Using a Network Approach to Examine Communication Patterns Within a Psychosis Social Support Group in Taiwan. *Journal of the American Society for Information Science and Technology*, 60, 7 (2009), 1504-1518.

[8] Civan, A., & Pratt, W. Threading Together Patient Expertise, *AMIA 2007 Symposium Proceedings* (Vol. 11). Chicago, IL (2007), 140-144.

[9] Cobb, S. Social Support as Moderator of Life Stress. *Psychomatic Medicine* 38, 5 (1976), 300-314.

[10] Coursaris, C.K., & Liu, M. 2009. An analysis of social support exchanges in online HIV/AIDS self-help groups. *Computers in Human Behavior* 25, 4 (2009), 911-918.

[11] Cunningham, J. A., T. van Mierlo, et al. An online support group for problem drinkers: AlcoholHelpCenter.net. *Patient Education and Counseling* 70, 2 (2008), 193-198.

[12] Cunningham, J.A, Humphreys, K., Koski-Jannes A., Providing personalized assessment feedback for problem drinking on the Internet: a pilot project. *J Stud Alcohol* 61 (2000), 794-8.

[13] Cunningham, J.A, Humphreys K., Kypri, K., van Mierlo T., Formative evaluation and the three-month followup of an online personalized assessment feedback intervention for problem drinkers. *J Med Internet Research* 8 (2006), e5.

[14] Cutrona, C.E., & Suhr, J.A. Controllability of Stressful Events and Satisfaction With Spouse Support Behaviors. *Communication Research*, 19, 2 (1992), 154-174.

[15] Eichhorn, K.C. Soliciting and Providing Social Support Over the Internet: An Investigation of Online Eating Disorder Support Groups. *Journal of Computer-Mediated Communication* 14, 1 (2008), 67-78.

[16] Himelboim, I. Reply distribution in online discussions: A comparative network analysis of political and health newsgroups. *Journal of Computer-Mediated Communication* 14, 1 (2008), 156-177.

[17] Humphreys K, Klaw E. Can targeting nondependent problem drinkers and providing internet-based services expand access to assistance for alcohol problems? A study of the moderation management self-help/mutual aid organization. *J Stud Alcohol* 62 (2001), 528-32.

[18] Kielstra, P. Doctor innovation: Shaking up the health system. (June 19, 2009) http://www.pewinternet.org/Media-Mentions/2009/Doctor-innovation-Shaking-up-the-health-system.aspx

[19] King, S. Analysis of Electronic Support Groups for Recovering Addicts. *Interpersonal Computing and Technology.* 2 (1994), 47-56.

[20] Mo, P. K. H., Malik, S. H., & Coulson, N. S. Gender differences in computer-mediated communication: A systematic literature review of online health-related support groups. *Patient Education and Counseling* 75, 1 (2009), 16-24.

[21] Pfeil, U. and P. Zaphiris 2007. Patterns of empathy in online communication. *Proc. of the SIGCHI conference on Human factors in computing systems*. San Jose, California, USA, ACM (2007), 919-928.

[22] Preece, J. 1999. Empathy online. *Virtual Reality* 4, 1 (1999), 74-84.

[23] Tanis, M. Health-Related On-Line Forums: What's the Big Attraction? *Journal of Health Communication: International Perspectives* 13, 7 (2008), 698 - 714.

[24] Wright, K.B. Computer-mediated support groups: An examination of relationships among social support, perceived stress, and coping strategies. *Communication Quarterly* 47, 4 (1999), 402-414.

Serving Library Users from Low-income Communities: Promoting Digital Literacy to eSociety

Yunfei Du, PhD
Assistant Professor
College of Information
Department of Library and
Information Sciences
University of North Texas

ABSTRACT

This study investigated the inclusion of low-income community users by surveying digital information literacy, in term of their attitudes toward electronic media and Internet-based reading. Two hundred thirty-eight public library users from ethnic diverse communities participated in this survey. Younger readers found reading online as easy as reading print books, while older library users preferred print media. Lower-income users were able to access the Internet mainly from libraries and reported slightly positively to online reading. Reading attitudes toward digital resources by residents of poorer urban communities did not vary by gender, contrary to common sense that women outread men. The implications of this result are explored; for example, being unemployed may allow underrepresented and minority users supplemental time for leisure reading. This study suggests approaches for iSchools and graduates to service diverse communities, promote digital information literacy, and bridge digital divide in information society.

Keywords

Digital literacy, online reading, information society, underserved users.

1. INTRODUCTION

Issues on eSociety drew attention from both professionals and researchers in recent years. For example, the International Conference on Information Society (http://www.i-society.eu/) sponsors an incoming track on eSociety which covers topics on social inclusion, intellectual property rights, computer-mediated communication, and social software.

One important area in eSociety research is the inclusion of underrepresented groups in information services, particularly promoting digital literacy to diverse users. Recent studies reported the influence of Internet to reading and literacy in the US. For example, the National Endowment of the Arts (2004, 2007) concluded that reading among the youth also dropped due to the possible influence of digital entertainment media. Several studies have reported similar trends globally. In the United Kingdom, annual book loans have been dropping since 1980, due mainly to the decrease in circulation of nonfiction and adult fiction (Grindlay and Morris, 2004). The widespread use of computers and of the Internet is likely a factor of the change of reading behavior (p. 609), considering the Internet facilitates reading as both a leisure activity and a source of information, and has become an alternative to print reading in the industrialized world.

While past research discussed how libraries championed and helped create digital resources to give users convenient access from a plurality of consoles and whereabouts (Bertot and McClure, 1999; McClure, Bertot, and Beachborad, 1996), few studies examined the changes in online reading by users from low-income communities. This study investigated how users from diverse demographic background reacted to ease of reading on the Internet. Research related to this topic can enhance knowledge of digital reading literacy, help to understand how information society evolve, and provide guidance on inclusion of underserved population to information society. This paper examined the concept of "ease of online reading," which is defined as how accessible or inaccessible electronic-based text is to the reader. The following are the research questions:

1. Are library users' perceptions of ease of online reading affected by demographic differences within low-income communities?

2. What is the influence of digital media on reading literacy in low-income communities?

2. REVIEW OF RELATED RESEARCH

2.1 Information Services to Diverse Communities

Reading literacy studies traditionally focused on reading cultures; while the core values of reference and information services include reader's advisories and providing users a variety of feedback vehicles (Pawley, 2002; Wiegand, 1998). Previous research by librarians and library scholars indicated that public libraries helped users from low-income communities experience their initial and ongoing access to electronic resources. Libraries also fostered reading and provided community networking opportunities on site such as storytelling periods for children, book fairs, and author readings. Chatman (1985a) found low-income library users to be relatively low consumers of television and, therefore, to prefer print media. They perceived print as the most credible and television as the least credible of mass media formats. Pettigrew, Durrance, and Vakkari (1999) reviewed the role of public libraries in community information services and urged for increased research on how networked community information influences citizens' daily lives, and how it affects their overall information-seeking behavior. Bishop and Bauer (2002) investigated what strategies (such as providing young

adults more food to eat and giving them inviting surroundings) and programs (such as an overnight slumber event: "Library Survivor") bring child and teenage readers to public libraries. The authors identified availability of the Internet, opportunities to volunteer, and capability to support academic research as libraries' most important attractants to users of youth (p. 42).

Bertot, McClure, and Ryan (2002) summarized the importance of implementing ideal technologies and training library staff to be able to assist patrons, so that public libraries can continue to function as key institutions for diminishing the digital divide. Accessing technology is only a first and relatively low-level step in the information literacy process. Once individuals gain access, they need then minimally to comprehend how to navigate the content in order to locate, retrieve, and evaluate useful information, and to synthesize this information in order to solve their information problem (Bertot, 2003). The digital divide was described by Bertot as multidimensional and complex, and extends beyond access to technology.

Researchers also reported strategies to engage people from lower-income communities in reading. Usherwood and Toyne (2002) surveyed both users and nonusers of public libraries and found that reading literature is a special activity that satisfies certain needs: escaping from reality, relaxing, bringing in knowledge, and assisting with personal development. McLoughlin and Morris (2004) examined the role of public libraries in advancing reading in adults having poor literacy, in a case study in the United Kingdom. They summarized strategies such as the use of reading groups, audio books, themed activities and events, and partnering with other libraries to offer rooms and spaces for "Pleasure of Reading" courses originating from local community colleges (p. 42). For example, Krashen and Shin (2004) detected that children from high-income families read more over the summer because of their access not only to public libraries, but also, in cases, to school and university libraries, as well as to their own or to their parents' home bookshelves. They discovered that summer after summer, poorer children did not have that opportunity and, over several years, fell behind in reading level. They suggested that public libraries proactively invite families to the library during periods of school adjournment, as well as increase their collections and tailor their summer hours to keep their services available to children of poverty. Williams (2005) related her experience on serving the financially challenged in Columbus, Ohio, particularly "Spontaneous Reading," in which librarians approach children or children do librarians, resulting in a librarian reading a book to children or even to one child. "Spontaneous Reading" attempts to be a model to families, showing parents and guardians how to engage in dialog with their children about books, all to help break the cycle of illiteracy and poverty.

2.2 Digital Media and Reading Literacy

According to the transaction theory, a person interacts with reading content like a river connects with its banks, each working its effects upon the other (Rosenblatt, 1986, 1994; Rosenberg, 1996). Digital media are different from print reading materials. McEneaney (2006) stated that users picture online documents as networks of nodes and links. This requires that readers define text structure by choosing links, which are based on readers' internal knowledge structure rather than on an author-defined text-structure (McEneaney, 2003, 2006; Rosenberg, 1996).

Digital media, identified by the National Endowment for the Arts (2004) as "TV, Internet, and computer games," have been recognized as important factors impacting leisure reading. Voluminous research points to how TV influences reading (Salomon, 1979; Reinking, 2001; van der Voort, 2001). Chatman (1985b) discovered that users from low-income communities usually do not have time for leisure activities, with reading being rated as one of the top leisure activities. She found low-income users' time for avocations, such as reading or visiting the library, usually coincided with the day's end, after other obligations had been met. Chatman and Pendleton (1995) disqualified mass media (TV and newspapers) as relevant information sources, considering them instead as sources of recreation, mere "escape and diversion" instead of "information" for the poor (p. 137). They concluded that economically disadvantaged members of society have a gap in second-level knowledge, or "knowledge about that which they do not know" firsthand, or are able to relate directly to personal or local circumstance (p. 143).

While watching TV and playing computer games might often amount to "learning," each may have different purpose, procedure, or cognitive effect other than reading, of either Internet or print reading. Hughes-Hassell and Lutz (2006, p.41) reported that middle school students who do not enjoy reading would rather play video games (44%) or watch TV (56%). This may be because students today are technology savvy, spending significant hours surfing the Internet, watching TV, and playing video games. They are able to multi-task among several Web sites and technologies; browse online, search for information, chat, and email, while using the same computer to do homework, talking and/or text messaging or playing games with their digital phones. They also expect to receive information quickly and efficiently (Lacina, 2005, p. 119). The digital age has led to changes in how young people think, learn, give, receive and create information, and how they interact with resources. To accommodate such changes, authors, illustrators, editors, designers, and publishers have been producing books that integrate the digital-age characteristics of interactivity, connectivity, and accessibility (Dresang, 2002; Dresang and McClelland, 1999).

The Internet may already have changed readers' attitudes towards presentation and format. Recently unveiled formats, such as online episodes/chapters of Internet-published books, or PDF files of articles and books, may have accustomed readers to viewing pages electronically. Also, the Internet is an operative tool for searching for information; however, concern is growing that the Internet might isolate people socially, and that youths may opt to converse electronically with their friends and to surf, rather than to use the Internet for reading. Schmar-Dobler (2003) found readers applied to their Internet reading previously-adopted

strategies for reading print text; at web pages one might read the brightened and bold headings and sought the topic sentences of paragraphs as a strategy to determine information applicability and whether to read on for completion. She further recognized that when reading on the Internet "guiding questions must be in the forefront of the reader's mind" or readers becoming "lost or sidetracked is likely" (p. 84).

However, while the Internet bridges the gap of the digital divide, recent studies signify that excessive use of the Internet might reduce work efficiency. Using data from a national random sample of American adults (N = 4,113), Nie and Erbring (2002) discovered that the more time people spent using the Internet, the more they lost contact with their social environment and the more they watched TV. Digital media may have changed regular users' reading behaviors by increasing "browsing/scanning," increasing "on-time reading," and decreasing "in-depth," "sustained" reading (Liu, 2005).

Understanding how economically-disadvantaged patrons use the Internet and whether digital media, especially the Internet, impacts their leisure reading, is critical in advancing literacy; and it is literacy that best generates a pathway for an individual to rise out of poverty. Such digital literacy may help to counteract the disparities in the Internet use among different social-economic groups (Gui, 2007). Once a low-income resident gains competence in reading, he or she can then decipher food labels, follow written instructions and precautions, and write responses on job and college applications. Then, once having become strengthened readers, low-income residents' can enjoy more promising futures which could include careers or better-paying occupations; enter into training programs or colleges; broaden knowledge of community, municipality, nation, and world; create or continue businesses; and practice reading to their children or elders. Ross (2003) urged researchers to explore how readers actually engage in different media, the reasons they choose one format over another, and their preferences among formats. Findings from such studies may help to secure public library funding for reading materials and for promote reading literacy to diverse and/or underprivileged communities. More reviews of literature related to reading and literacy can be located in works by Radway (1994), Ross (1999, 2003, 2005), Bawden (2001), Liu (2005), Mackey (2007), and Du (2009).

3. METHOD

Prior to this study, the author has conducted two preliminary studies to refine the questionnaire used in this study. The questionnaire allows respondents to rate each question item using Likert scale (1= strongly disagree, 5= strongly agree). The scale was created by Likert (Babbie, 2001) and is commonly treated as an interval level measure in the social sciences (Gross & Saxton, 2006). The questionnaire generated a moderate reliable score in previous studies (Cronbach alpha = .748).

This study surveyed library patrons from four of the twenty-three branches of the in a mid-west urban library and the central branch of a suburban library. Among the four public library branches, one in the northwestern city had mostly Hispanic and Middle Eastern users, while three in the southern and eastern regions hosted African-American communities. The suburban library had more low-income European Americans. The study's limitation was its stratified convenience sampling rather than random sampling. The date of library visits and their locations were chosen with the help of library administrators to best represent multiple ethnic groups. Research assistants distributed questionnaires directly to library patrons in library buildings.

After collecting the data, the research team calculated descriptive statistics and conducted analysis of variances (ANOVA) to see how different grouping information, such as age and income level influenced participants' attitudes toward reading online. The dependent variable was library patrons' perception of online reading difficulty, and the independent variables were age groups, and income levels. ANOVA tests examined four null hypotheses:

H1: There is no statistically significant difference in library patrons' perception of ease of online reading among various age groups.

H2: There is no statistically significant difference in library patrons' perception of ease of online reading among various income levels.

H3: There is no statistically significant difference in library patrons' perception of ease of online reading by difference in gender.

H4: There is no statistically significant difference in library patrons' perception of ease of online reading among various ethnicities.

4. RESULTS

Two hundred thirty-eight library patrons completed the questionnaire during their library visits. Among them, 87 (38%) were male and 141 (62%) were female. Twenty-eight percent were 18 to 25 years, 24% were 26 to 35; 16% were 36 to 45, and 19% were 46 to 55. Only 13% were 56 or older. Among the participants, 167 (73.2%) claimed to be urban, and 26.8% claimed to be suburban or rural, albeit probably all should be considered "urban" as the term is broadly defined. Fifty-six percent (or 124) of the participants were African American, 22% (51) Caucasian, 14% (31) Hispanic, and only 6% (17) interracial or other. The surveyed area possessed a relatively high unemployment rate. Hence, this study surveyed income level, indicated by average annual incomes of the participants of those neighborhoods. Twenty-eight percent of the people did not report. Twenty-one percent were living $10,001 or below, 28% between $10,001 and $30,000, and 29% between $30,001 and $50,000. The $50,001 or above accounted for only another 22%. Most participants commented that they did not have a computer at home, and used library computers to access the Internet.

Table 1 illustrates descriptive statistics of answers to the first 13 questions asked in the Appendix. Answers were clustered based on mean scores (1 = strongly disagree, and 5 = strongly agree). Positive answers include Questions 6 (how video impacts reading), 10 (parents' influence), 11 (volunteering in libraries), and 12 (summer reading programs). Negative answers include Questions 2 (Internet

browsing), 4 (email), and 7 (too many books). Neutral answers include Questions 1 (TV), 3 (Internet to find books), 5 (chat), 8 (movies), 9 (sports), and 13 (reading online).

Table 1. The Influence of Digital Media on Reading Literacy

Questions	Mean	SD	Skewness	Kurtosis
Positive Answers				
A6. Video impacts reading	3.91	1.19	-.96	.05
A10. Parents' influence	4.01	1.25	-1.2	.37
A11. Volunteering in libraries	3.86	1.07	-.70	-.08
A12. Summer reading programs	4.2	1.01	-1.45	2.0
Negative Answers				
A2. Internet browsing impacts reading	2.51	1.33	-.01	-1.38
A4. Using e-mail impacts reading	2.52	1.36	.48	-1.02
A7. Too many books to choose from	2.28	1.30	.73	-.59
Neutral Answers				
A1. Watching TV impacts reading	3.01	1.44	-.01	-1.38
A3. Using the Internet to find books	2.58	1.30	.41	-.90
A5. Using Chat impacts reading	2.77	1.45	.17	-1.32
A8. Watching movies impacts reading	2.95	1.33	.11	-1.05
A9. Sports impact reading	2.97	1.17	-.02	-.71
A13. Reading online same as print books	2.95	1.33	.11	-1.05

Note: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree

In order to answer Question 1, one-way Analysis of Variance (ANOVA) was conducted to estimate the effects of Gender and Income Level on Ease of Online Reading. Table 2 displays the ANOVA results.

Table 2. Ease of Online Reading by Age Groups

	Sum of Squares	df	MS	Sig.	eta^2
Between Groups	28.69	5	5.74	.01*	.07
Within Groups	364.68	220	1.66		
Total	393.34	225			

Note: Statistically significant at .01 level. Levene's Test of Homogeneity was statistically nonsignificant at .93, which secures the use of ANOVA.

From Table 2, the ANOVA yielded a statistical significance: a p-value of .01. ANOVA is a statistical technique to estimate the difference between means of groups. A statistically significant p-value indicates statistically significant differences among the groups. Because statistically significant difference is influenced by sample sizes, effect sizes should be reported. In this study, the effect size is .07 in terms of Eta square. Thus one can reject the null hypothesis and substantiate the research question that a statistically significant difference in library patrons' perceptions of ease of online reading exists across different age groups.

In order to find out what contributed to the difference, the author conducted the least significant difference (LSD) post hoc tests. The LSD t-test coefficients identified that the younger group (18 to 24) was statistically significantly different from all other age groups: 25 to 34 (α = .01), 35 to 44 (α = .01), 45 to 54 (α = .02), 55 to 64 (α = .01), and 65 or older (alpha = .02). The means for age groups were illustrated in Figure 1.

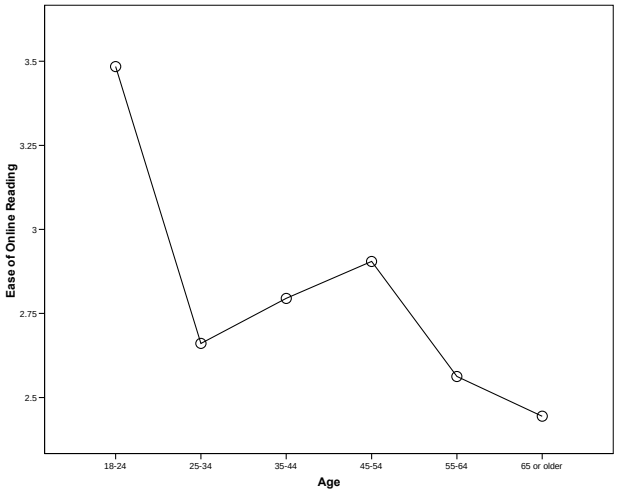


Figure 1. Ease of Online Reading by Age Differences

From Figure 1 above, participants from 18 to 24 tend to agree that reading online is as easy as reading print books, while the other three groups' opinions were either neutral or negative. As reported above, the difference is statistically significant.

In order to answer research Question 2, the author conducted another ANOVA to see how people with different income levels reacted differently to Ease of Online Reading. The survey did not directly ask the individual's actual annual income due to local regulations, but indirectly sought the average income of the average household of the neighborhood where he or she resided. The limitation is minimized because historically people with same social economical status are clustered together in this city and each branch library serves residents in their own communities. Table 3 displays the ANOVA results.

Table 3. Ease of Online Reading by Income Level

	Sum of Squares	df	MS	Sig.	eta^2
Between Groups	24.65	3	8.22	.00*	.08
Within Groups	269.62	163	1.65		
Total	294.28	166			

Note: Statistically significant at .01 level. Levene's Test of Homogeneity was statistically nonsignificant at .87, which secures the use of ANOVA.

From Table 3, the ANOVA yielded statistical significance with p-value <.01. Thus one can reject the null hypothesis and support the research hypothesis that there is a statistical

significant difference on library patrons' perception of ease of online reading among those with different income levels.

In order to find out what contributed to the difference, the author conducted LSD post hoc tests. The statistics identified that the more affluent group ($50,001 or above in family income) was statistically different from other groups: $10,000 or less (α = .00), $10,001 to 30,000 (α = .02), and $30,001 to 50,000 (α = .00). The means for all age groups are illustrated in Figure 2.

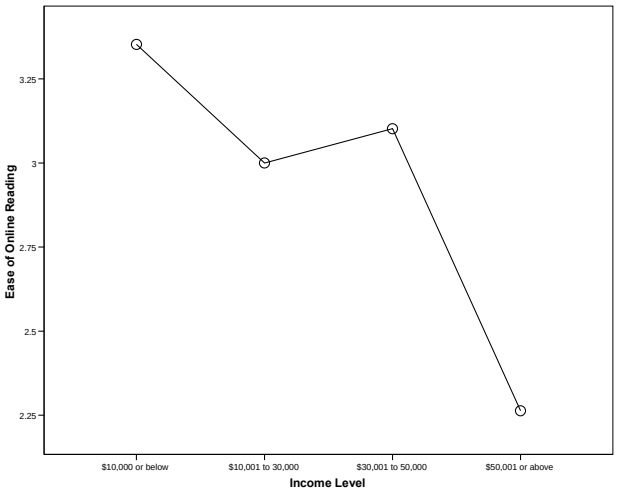


Figure 2. Ease of Online Reading by Income Level

ANOVA tests on hypothesis 3 and 4 yielded statistical non-significant results. Thus, the author rejected the null hypothesis. Contrary to common sense that women outread men, the data from current sample did not find any statistical differences among males and females, nor differences among different ethnic groups, on their perceptions on reading online. It seems urban men and women have similar potential online resources.

5. Conclusions and Implications

This study confirms that libraries hold tremendous importance in promoting digital literacy to diverse population. Patrons from low-income communities are often only able to access the Internet through public library outlets. They may not have the financial resources to set up a computer at home and install Internet service. Also, their using the Internet has proven not so excessive that it impacts their either online or print reading time. Using the Internet at a public library remains the principal means for a low-income citizen to achieve on-line access.

Furthermore, lower-income residents, because of their limited access to the Internet, might treat using the Internet as a privilege equal to paper media, hence they enjoy reading online and may not initially realize that all the Internet offers includes activities and entertainments other than reading, and that, one day for them, the Internet may compete for their leisure-reading time. Higher-income residents, who might read and work over the Internet on a daily basis, might more readily enjoy print books as a luxurious way to spend leisure time.

Traditionally, libraries have designed digital resources to meet to the needs of younger, technology-savvy user groups. The findings in this study also suggested involve all age groups in information services to meet the needs of information society. As Hagood (2003) suggested, "new media and online illiteracies belong to and affect people of all ages" (p. 23).

Liu (2006) ascertained that of graduate students having come to an academic library to complete their assignments and essays, 51.9% turned first to online library information resources (e.g., e-journals) and 28.6% to the Web. His findings about students in collegiate milieus parallel those about younger users in public library settings. In addition, users belonging to older adult generations tend to prefer print rather than digital resources.

Based on this sample, library patrons tended to agree that browsing the Internet or checking email did not influence leisure-time reading, while other digital media such as watching TV, or playing computer games seemed to compete with their reading time. It seems users from different reader groups react differently to print and electronic reading materials. Younger generations and low-income users read over the Internet as comfortably as they read print formats. A follow-study might identify what led to each individuals like or dislike online reading, and whether lower income was caused by younger age. Future efforts seem vital to advocate digital literacy to diverse communities, since these endeavors advance reading and literacy, and the attainment of literacy skills can lead underserved populations toward training, education, and employment.

Currently library and information science (LIS) students may choose to purse online degrees quickly while distancing themselves from users with diverse background. To promote digital literacy to diverse communities, iSchool educators should prepare LIS students to be able to communicate with diverse audience, teach information literacy to users with different background, be sensitive to needs of their communities, even further, and engage in global social information exchange. Library and information agencies still face the challenging job of supplying equal access and service to diverse patrons - women, youth, the aging, people with disabilities, racial and ethnic minority readers, and user groups from different income levels to overcome the digital divide. More support to iSchools education is needed to create workforce to unemployed, underserved urban citizens to acquire knowledge and skill online, transform the daily living of underserved populations, and ensure digital equality.

6. REFERENCES

[1] Babbie, E. (2001). The practice of social research (9th ed.). Belmont, CA: Wadsworth/Thompson Learning.

[2] Bawden, D. (2001). Information and Digital Literacies: A Review of Concepts. Journal of Documentation 57(2), 218-259.

[3] Bertot, J. C. (2003). The Multiple Dimensions of the Digital Divide: More Than the Technology "Haves" and "Have Nots." Government Information Quarterly 20(2), 185-191.

[4] Bertot, J. C., & McClure, C. R. (1999). U. S. Public Library Outlet Internet Connectivity: Progress Issues and Strategies. Library and Information Science Research 21(3), 281-298.

[5] Bertot, J. C., McClure, J. R., & Ryan, J. (2002). Impact of External Technology Funding Programs for Public Libraries: A Study of LSTS, E-Rate, Gates, and Others. Public Libraries 41(3), 166-171.

[6] Bishop, K., & Bauer, P. (2002). Attracting Young Adults to Public Libraries: Frances Henne/YALSA/VOYA Research Grant Results. Journal of Youth Services in Libraries 15(2), 36-44.

[7] Chatman, E. A. (1985a). Low Income and Leisure: Implications for Public Library Use. Public Libraries 24(1), 34-36.

[8] Chatman, E. A. (1985b). Information, Mass Media Use and the Working Poor. Library and Information Science Research 7(2), 97-113.

[9] Chatman, E. A., & Pendleton, V. E. M. (1995). Knowledge Gap, Information-Seeking and the Poor. The Reference Librarian 23(49-50), 135-145.

[10] Du, Y. (2009). Librarians' Responses to "Reading at Risk": A Delphi Study. Library & Information Science Research, 31(1), 46-53.

[11] Dresang, E. (2002). Harry Potter and Censorship. Florida Media Quarterly, 27(4), 8-11.

[12] Dresang, E. T., & McClelland, K. (1999). Radical Change: Digital Age Literature and Learning. Theory into Practice 38(3), 160-167.

[13] Grindlay, D. J. C., & Morris, A. (2004). The Decline in Adult Book Lending in U.K. Public Libraries and its Possible Causes. Journal of Documentation, 60(6), 609-631.

[14] Gross, M., Saxton, M. L (2002). Integrating the imposed query into the evaluation of reference service: A dichotomous analysis of user ratings. Library & Information Science Research, 24(3), 251-263

[15] Gui, M. (2007). Formal and Substantial Internet Information Skills: The Role of Socio-Demographic Differences on the Possession of Different Components of Digital Literacy. First Monday 12(9). Available at: www.firstmonday.org/issues/issue12_9/gui/index.html (accessed November 17, 2009).

[16] Hagood, M. C. (2003). New Media and Online Literacies: No Age Left Behind. Reading Research Quarterly 38(3), 387-391.

[17] Hughes-Hassell, S., & Lutz, C. (2006). What Do You Want to Tell Us About Reading? A Survey of the Habits and Attitudes of Urban Middle School Students toward Leisure Reading. Young Adult Library Services 4(2), 39-45.

[18] Krashen, S., & Shin, F. (2004). Summer Reading and the Potential Contribution of the Public Library in Improving Reading for Children of Poverty. Public Library Quarterly 23(3/4), 99-109.

[19] Lacina, J. (2005). Media Literacy and Learning. Childhood Education 82(2), 118.

[20] Liu, Z. (2005). Reading Behavior in the Digital Environment: Changes in Reading Behavior over the Past Ten Years. Journal of Documentation 61(6), 700-712.

[21] Liu, Z. (2006). Print vs. Electronic Resources: A Study of User Perceptions, Preferences, and Use. Information Processing and Management 42(2), 583-592.

[22] Mackey, M. (2007). Literacies across Media: Playing the Text, 2nd edition. New York: Routledge.

[23] McClure, C. R., Bertot, J. C., & Beachboard, J. C. (1996). Enhancing the Role of Public Libraries in the National Information Infrastructure. Public Libraries 35(4), 232-238.

[24] McEneaney, J. E. (2003). A Transactional Theory of Hypertext Structure. In 52nd Yearbook of the National Reading Conference, eds. Fairbanks, C. M., Worthy, J., Maloch, B., Hoffman, J. V., & Schalert, D. L., 272-284. Oak Creek, W.I.: National Reading Conference.

[25] McEneaney, J. E. (2006). Agent-Based Literacy Theory. Reading Research Quarterly 41(3), 352-371.

[26] McLoughlin, C., & Morris, A. (2004). U.K. Public Libraries: Roles in Adult Literacy Provision. Journal of Librarianship and Information Science 36(1), 37-46.

[27] National Endowment for the Arts (2004). Reading at Risk: A Survey of Literary Reading in America. Available at: www.nea.gov/pub/ReadingAtRisk.pdf (accessed November 17, 2009)

[28] National Endowment for the Arts (2007). To Read or Not To Read: A Question of National Consequence, Available at: www.nea.gov/research/ToRead.pdf (accessed November 17, 2009)

[29] Nie, N. H., & Erbring, L. (2002). Internet and Society: A Preliminary Report. IT and Society 1(1), 275-283.

[30] Pawley, C. (2002). Seeking "Significance": Actual Readers, Specific Reading Communities. Book History 5, 143-160.

[31] Pettigrew, K. E., Durrance, J. C., & Vakkari, P. (1999). Approaches to Studying Public Library Networked Community Information Initiatives: A Review of the Literature and Overview of a Current Study. Library and Information Science Research 21(3), 327-360.

[32] Radway, J. (1994). Beyond Mary Bailey and Old Maid Librarians: Reimagining Readers and Rethinking Reading. Journal of Education for Library and Information Science 35(4), 275-296.

[33] Reinking, D. (2001). Multimedia and Engaged Reading in a Digital World. In Literacy and Motivation: Reading Engagement in Individuals and Group, eds. Verhoeven, L., & Snow, C. S., 195-221. Mahwah, N.J.: Lawrence Erlbaum.

[34] Rosenberg, J. (1996). The Structure of Hypertext Activity. In Proceedings of the Seventh Hypertext Conference of the Association for Computing Machinery, Bethesda, MD, 22-30. New York: Association for Computing Machinery. DOI= http://doi.acm.org/10.1145/234828.234831

[35] Rosenblatt, L. (1986). The Aesthetic Transaction. Journal of Aesthetic Education 20(4), 122-128.

[36] Rosenblatt, L. (1994). The Reader, the Text, the Poem. Carbondale: Southern Illinois University Press.

[37] Ross, C. S. (1999). Finding without Seeking: The Information Encounter in the Context of Reading for Pleasure. Information Processing and Management 35(6), 783-799.

[38] Ross, C. S. (2003). Reading in a Digital Age. Available at: http://66.102.1.104/scholar?hl=en&lr=&q=cache:-4e--YaUMbAJ:www.camls.org/ce/ross.pdf+c+s+ross+reading+in+a+digital+age (accessed November 17, 2009)

[39] Ross, C. S. (2005). Reader Response Theory. In Theories of Information Behavior, eds. Fisher, K. E., Erdelez, S., & McKechinie, L., 303-307. Medford, N.J.: Information Today.

[40] Salomon, G. (1979). Interaction of Media, Cognition, and Learning. San Francisco: Jossey-Bass.

[41] Schmar-Dobler, E. (2003). Reading on the Internet: The Link between Literacy and Technology. Journal of Adolescent and Adult Literacy 47(1), 80.

[42] Usherwood, B., & Toyne, J. (2002). The Value and Impact of Reading Imaginative Literature. Journal of Librarianship and Information Science 34(1), 33-41.

[43] van der Voort, T. H. A. (2001). Television's Impact on Children's Leisure-Time Reading and Reading Skills. In Literacy and Motivation: Reading Engagement in Individuals and Groups, eds. Verhoeven, L., & Snow, C. E., 95-119. Mahwah, N.J.: Lawrence Erlbaum.

[44] Wiegand, W. A. (1998). Introduction: Theoretical Foundations for Analyzing Print Cultures as Agency and Practice in a Diverse Modern America. In Print Culture in a Diverse America, eds. Danky, J. P., & Wiegand, W. A., 1-13. Chicago: University of Illinois Press.

[45] Williams, C. (2005). Serving Urban Populations. Public Libraries 44(6), 321.

Mining for Culture: Reaching Out of Range

Wanda Eugene
Auburn University
3101 Shelby Center for Eng. Tech.
Auburn, AL 36849-5347 U.S.A.
eugenwa@auburn.edu

Juan E. Gilbert, Ph.D.
Clemson University
821 McMillan Rd.
Clemson, SC 29634-0974 U.S.A.
juan@clemson.edu

ABSTRACT
The goal of this paper is to present a tool that will sustain the development of culturally relevant computing artifacts by providing an effective means of detecting culture identities and cultures of participation. Culturally relevant designs rely heavily on how culture impacts design and though the guidelines for producing culturally relevant objects provide a mechanism for incorporating culture in the design, there still requires an effective method for garnering and identifying said cultures that reflects a holistic view of the target audience. This tool presents culturally relevant designs as a process of communicating with key audiences and thus bridging people and technology in a way that once seemed out of range.

Categories and Subject Descriptors
H.2.8 [**Database Applications**]: Data mining

General Terms
Design

Keywords
Data Mining, Ethnocomputing, Culture, Design

1. INTRODUCTION
Although culture has made its way to the forefront of conversation, constantly a caveat to be accounted for, it still manages to maintain its chameleon image. In every context of its use, it takes on a different meaning or representation, fulfilling a different purpose or function. This lack of uniformity in its functional meaning has become less of a concern than its apparent impact on our changing global society. Yet and still, its increasingly significant presence requires its inclusion in all present and future innovations. Thomas Hughes reflected on the idea that since the beginning of the 20th century, all inventions are fashioned by individuals with a very specific educational and cultural background [10]. He explained that each part of an invention's complex story involves processes that are highly contingent and highly intertwined with social, economic, and political relationships.

Culture and culture identities come from somewhere, have histories, and undergo constant transformation [9]. It can also be said that the nuances of a culture are best understood by its participants. As culture influences action by shaping a repertoire or "tool kit" of habits, skills, and styles from which people construct "strategies of action" [15], it creates culture identities, which are understood within and between the culture participants. These tool kits depict the wealth of knowledge shared and understood within the culture. They also provide an alternative lens of understanding and interpreting data not already associated with one's mental schema. It is suggested that the interaction of culture, affect, and cognition allows a person to develop multiple intelligences, interpersonal intuition and deep knowledge of oneself [11]. However it is also acknowledged that people are unique individuals who belong to several different identity groups [14]. With attempts to impact a broader audience, our efforts converge upon a means to capture a better understanding on these unique culture identities and to bridge people and technology in a way that once seemed out of range.

As of late, there has been several research endeavors purposed for the task of appropriating culture for use in technological design. This paper discusses culture in the context of the design and development of computing technology and further suggests a means for doing culture discovery – the Culture Inquiry Form, which is then introduced and explained. The goal of the Culture Inquiry Form is to present a process that draws upon existing computer science tools such as data mining, and attempts to emulate known methods such as ethnography research, all to serve as a means for better depicting cultures of participation and culture identity.

2. RELATED WORKS
2.1 Culture in the Field
Numerous researchers have studied the use of culture to relate a variety of concepts to diverse learners and to understand subject specific concepts among culture groups. For example, computing as an element within culture can be found in various artifacts and practices within a community of practice. Eglash investigates fractal geometry as in geometric patterns, calculations, and theories, as facets expressed in various African cultures [3]. The use of culture to make connections has been used in several disciplines and domains. For example, Carol Lee uses culture modeling to teach literature. Culture modeling, in essence, provides "instructional organization that makes academic concepts, strategies, and habits of mind explicit that makes ways of engaging in the work of the disciplines familiar and that provides supports for instances where the learner is unsure" [13]. She uses the culture of everyday practices as a lens for understanding the role of perception in influencing actions. Within culture modeling, culture data sets are used, which are

familiar examples that new learning can be anchored and used to provide problems whose solutions mirror the demands of the academic task that we want the learner to discover [13]. Making connections across relevant schemata or clusters of schematic networks helps to create connections between the known and the unknown.

2.2 Cultural Data Mining

Data mining provides a means of transforming large groups of data into information by extracting a pattern and also designates fitting a model to data, finding structure from data, or in general, any high-level description of a set of data [5]. Data mining algorithms' ability to extract patterns from data facilitates a growing need to analyze a subset of data or a model applicable to that subset, within a large data set. As we quickly move from data sets consisting of kilobytes to now terabytes of data, such as that used by the Library of Congress or Wal-Mart, it quickly becomes a daunting task to extract useful information. As computers grow in speed, number-crunching capabilities, and memory, scientific researchers are edging into data overload as they try to find meaningful ways to interpret these data sets [12]. Thinking of the notion of the varying culture identities that exist in our society, data mining offers a means of extracting these unique patterns enumerated from data, as appose to relying upon assumption or sweeping generalizations. For this reason, the idea of cultural data mining is taking root. Manovich observes that until now, the study of cultural processes relied on two types of data: shallow data about many people (statistics, sociology) or deep data about a few people (psychology, ethnography, etc) [2]. Consequently, we can now collect detailed data about very large numbers of people, objects and/or cultural processes and we no longer have to choose between size and depth [2].

2.3 Design of Cultural Relevant Software

Young's Culture Based Model (CBM) reflects a model of culture that evolved from historical and linguistic analyses, in which the findings extrapolated from the analyses reveal a treasure of cultural remnants [16]. The cultural remnants provide an intercultural instructional design framework that guides designers through the management, design, development, and assessment process, while taking into account explicit culture-based considerations [4]. Young's model poses high-level questions to facilitate the big picture of the management of undertaking the design process. However, designers stemming from a technical background, such as computer science, have found this model difficult to navigate for software designers in need of a direct guide to support the design and evaluation of a software artifact.

Cultural Relevance Design Framework assists designers or design teams with creating culturally authentic technology [4]. This framework is designed to uncover the design team's beliefs and biases about their target audience, highlight aspects of about the target audience that might be unknown, and suggest cultural assets that can be investigated to provide building blocks for sound cultural representations [4]. The authors define culture within two dimensions, presenting a wide range of attributes that can be compiled to further capture and illustrate the concept of culture. To guide the design of culturally relevant tools [4] depicted these two dimensions within four themes: Practices, Ontology, Representation, and Tasks. The Cultural Relevance Design Framework is organized such that each of the themes are presented with a definition, an investigative question, and suggested criteria to help the designer explore and better understand the culture of the target audience. The framework provides concrete criteria that correlate to the socio-cultural norms of the targeted group of users [4]. Overall, this framework informs decisions regarding cultural relevance at the onset of the design process as well as a method of evaluating the cultural relevance throughout production processes to help ensure that the goal of a culturally relevant design is produced [4].

3. CULTURE INQUIRY FORM

Much research has focused on the need of culture and more recently how to use culture, but limited research has been proposed on a feasible means for capturing culture in a useful way that can be easily incorporated in design and development. Though the models discussed above address the much-needed frameworks and guidelines to effectively engage in designing culturally authentic technology, still absent from this discussion however, are how to obtain a persons culture identity information. Lee's culture modeling and Eglash's discovery of African fractals came as a result of extensive ethnographic research, that is normally not afforded to a designer engaged in software development. So the challenge becomes how to obtain information, which entails a holistic view of people, where the cultural preferences and differences of the target audience that should influence the design and development of a technology are accessible? Therefore, we present the Culture Inquiry Form (CIF) to serve as an intake of the target audience cultures of participation, and a snapshot of their culture identity.

3.1 CIF

The Culture Inquiry Form (CIF) allows the learner to self identify with the culture(s) in which they participate. CIF collects culture participation information based on "who you are" and "what you do" [8]. The first part, or the "who you are" portion, includes the demographics section of the instrument, which was designed to correlate the data collecting techniques of the U.S. Census Bureau and the Department of Labor, giving a consistent means of measurement. The US Census Bureau demographic categories serve as a model for this study, entailing questions such as age, ethnicity, and gender. The second part of CIF allow us to capture the "what you do" part of the learners' culture of participation. The design of the second part of CIF stems from the Culture Participation Focus Group Protocol and the collected data. The focus group was conducted in collaboration with the Information Management and System Engineering (IMSE) Program in Detroit, Michigan. The IMSE program, under Wayne State's NSF Broadening Participation in Computing Project, is a collaboration of Wayne State, Focus Hope, and several industry partners to support disadvantage students at critical junctures from a GED through the completion of a post secondary degree [1]. The focus group provided a means of understanding the cultures of participation of the targeted audience and their ontology for characterizing their participation in these cultures. The Culture Participation Focus Group Protocol

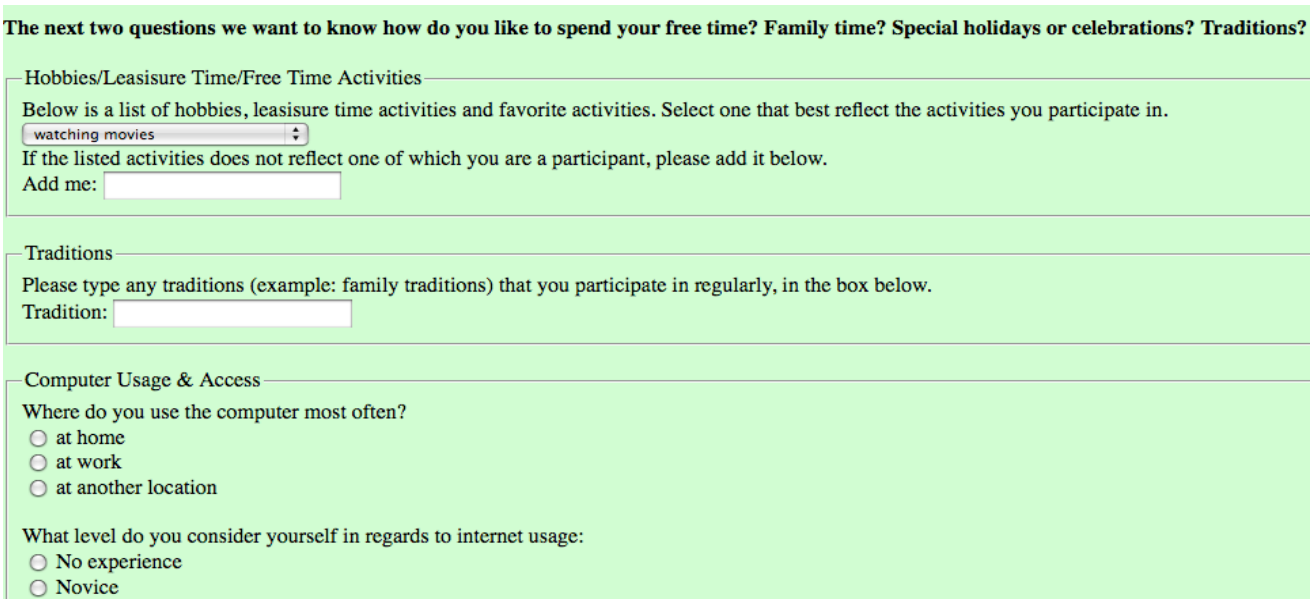


Figure 1. A Sample Portion of the Culture Inquiry Form.

was derived from the focus group protocol designed by the Family Math Team at Stanford University. Thus, the second part of CIF contains selections pertaining to hobbies, employment, and traditions. CIF also entails a field for learners to further describe their participation in said culture(s). If the learner is unable to identify with the listed hobbies another avenue is provided for learners to enter in the cultures in which they participate. A sample portion of CIF is displayed in Figure 1. Also included were questions pertaining to computer usage and perceived level of computer experience. CIF is designed to complement tools such as the Cultural Relevance Design Framework.

3.2 Applications Quest™

The data collected from the CIF is analyzed using cultural data mining, by running a clustering algorithm, Application Quest™, on this data to determine the dominant culture of participation among the participants. Applications Quest™ is a dynamic software tool developed to perform holistic comparisons using hierarchical clustering approach [6]. Applications Quest™ (AQ) takes in numerical values or nominal attributes to determine clusters of similar applications. AQ compares every application to every other application using n C r = n! / [(n-r)! r!], and places the result of each comparison into a database table called the similarity matrix. All numeric attributes are scaled to values between 0 and 1. When considering nominal values, the Nominal Population Metric (NPM) is used. The NPM begins by identifying the nominal attributes within the similarity matrix and then processes them as follows [7]:

1. Compute the total number of combinations for all applications using n C r.
2. Compute the number of unique nominal attribute values.
3. Compute the number of combinations for the unique nominal values using n C r.
4. For those combinations of the nominal attribute value pairs, compute the coverage percentage within the application similarity matrix.
5. The nominal population matrix shows nominal attribute pair coverage across all comparisons. This is an accurate measure of the impact of the nominal attribute value pairs based on their actual existence within the data population. The next step in this process is to adjust the Coverage values if necessary. This is the desired goal when the application is measuring difference vs. similarity.
6. The Coverage values in the nominal population matrix are now the Nominal Population Metrics that can be used in clustering algorithms to accurately compare nominal attribute values.

Using the squared Euclidean distance measure, AQ computes a similarity matrix. To determine the clusters, AQ uses a divisive clustering approach by identifying the two most different applications using the similarity matrix. Using the two most different applications, AQ forms clusters around them based on each individual application's closeness to one or the other.

4. CIF IN PRACTICE

To better illustrate CIF, and further explain the process of identifying cultures of participation, we present the findings of a study recently conducted.

4.1 Demographics

The participants were recruited from the student and faculty population of Auburn University (Auburn, AL). Overall, the study had 104 participants of whom 65% were male and 35% were female. 81% indicated that they were in the age range of 19-24, 16% said they were in the age range of 25-34, and 3% were in the age range of 45-54. Regarding ethnicity, 79% of participants identified themselves as Caucasian, 17% as African American, and 4% as Asian.

4.2 Findings & Analysis

In reviewing the collected data, an approach was selected which grouped attributes such as hobbies and traditions into buckets. All of the questions in CIF were designed as radio buttons or check boxes except for hobbies and traditions. The hobbies question presented the participant with a dropdown list of a several hobbies. These hobbies were those gathered from the focus group study. The participants were also permitted to enter their hobby if it did not appear in the provided list, and it would be added to the list that participants would see when CIF is loaded again. The traditions question was also designed as an entry field, so that participate could enter their traditions and describe them accordingly.

After reviewing the entries for hobbies and traditions, there were several overlaps, thus, it was decided to condense these separate entries into one larger grouping. For example, several entries included sports, football, basketball, and sports in general. All such entries were then identified under the larger category of sports. Another example, watching TV and watching movies, was condensed to the category of entertainment. A similar approach was taken for the traditions attribute. For the traditions attribute, the participants entered more of an explanation of what their tradition entailed. For example, one entry would be "Christmas dinner with the family". In analyzing the entries, we created buckets around the central themes and checked all that apply for each given tradition. Using the example above, holidays, dinner, and family would have been the buckets checked off. The majority of participants in this particular study didn't enter a tradition, and given the amount of variability in the entries, we didn't include it in the analysis. Application Quest™ was then run on the 104 collected responses as shown in the summary in Figure 2.

Upon uploading the data to AQ, the number of clusters (k) that we desired the responses to be grouped into must be determined. As with other clustering algorithms, there is no magic k. So we tried several numbers with the goal that the clusters would remain relevant, where irrelevant entries are not grouped or forced together, and that we don't have several clusters all having one entry. Thus, after several runs and trials, the number of clusters was set to nine.

AU AQ Application Summary (104 applications and 9 clusters)
Difference Index for all Applications 26.97% Standard Deviation 17
Difference Index for Recommended Applications 48.07% Standard Deviation of 14
Age: 19-24 (84), 25-34 (17), 45-54 (3)
CompLevel: intermediate (inter) (58), expert (34), novice (11)
CompUsage: home (89), work (8), another (6)
Education: some college (64), Diploma/GED (16), Graduate or professional (13)
Ethnicity: White (79), Black (18), Asian (4)
Frequency: Daily (96), Weekly (6), (1)
Gender: Male (68), Female (36)
Hobbies: Sports (46), Entertainment (22), Fishing (5)

Figure 2. AQ Summary

Of the nine clusters, 86 participants fell into clusters five, one, and eight. The dominant attributes of those clusters are depicted in Table 1.

Table 1. Table captions should be placed above the table

Cluster #	Number of Applications	Dominate Attributes		
5	44	Age, gender,	Usage, ethnicity, Frequency	
8	22	Usage, frequency	Gender, Level	Age
1	20	Frequency	Level	Gender

The clusters produced by Applications Quest™ provide an accurate and efficient way to determine the cultures of participation. The efforts of determining cultures of participation normally determined after extensive ethnography studies. Thus, AQ tremendously aids in the effort to capture the same essence of a learner in a quantitative approach. In this very basic study, it was easy to see and draw logical conclusions regarding each cluster. In running Application Quest™, industries and traditions were excluded in the specified attributes to simplify the data analysis. The similarities were so few they didn't contribute to the clusters. Given Table 1 along with the data, a designer is provided with fast access to knowledge of the cultures of participation of the participants that lay in the majority clusters.

5. CONCLUSION

The aim of this paper was to introduce a tool and a process for discovering cultures of participation and culture identity. This understanding aids with the inclusion of culture in the design and development of computing technology, such as culturally relevant products. The Culture Inquiry Form uses data mining instead of standard statistical data analysis to determine culture similarities. Standard statistical package can make it difficult to determine the cultures of participation, because they are often grouped based on frequency and give limited information regarding nominal attributes. Using statistics to create content, for example, if only given the information generated in Figure 2, it would not be as clear to detect a majority cluster profile with three similar attributes depicting the participants' culture of participation, especially in terms of large data sets such as culture. Applications Quest™ provides a means for identifying the cultures of participation based on attributes they have selected in common. The Culture Inquiry Form is presented cautiously, for is it not to be viewed as an effort to substitute ethnography research efforts, for it is on this very premise that we have access to the rich data and insight of humanity and culture in the deepest form. Instead, the Culture Inquiry Form is presented as an attempt to bridge the gap and bring together that rich knowledge of culture to the hands of the designers and developers to aid in developing computing technology that reflects the needs of our diverse society.

6. ACKNOWLEDGMENTS

This material is based in part upon work supported by the National Science Foundation under Grant Number CNS-0837580. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

7. REFERENCES

[1] Brockmeyer M. 2007. Broadening Participation in Computing at Wayne State University Retrieved May 20th, 2009. From http://www.cs.wayne.edu/~mab/BPC.html

[2] Digilib. 2009, June 22. Digital Humanities 2009: Opening Plenary by Lev Manovich Retrieved from

http://digilib.bu.edu/blogs/digilib/2009/06/dh09-opening-plenary-by-lev-manovich/

[3] Eglash. R. 1999. African Fractals: Modern Computing and Indigenous Design, Rutgers University Press, New Brunswick, NJ.

[4] Eugene, W. Hatley, L. McMullen, K. A. Brown, Q. Rankin, Y. Lewis, S. 2009. This is Who I am and This is What I do: Demystifying the Process of Designing Culturally Authentic Technology. In HCI International 2009 13th International Conference on Human-Computer Interaction, San Diego, California, July 2009

[5] Fayyad, U., Shapiro. G. P., and Smyth, P. 1996. Knowledge Discovery and Data Mining: Towards a Unifying Framework. KDD-96, 82-88.

[6] Gilbert, J.E. 2006. Applications Quest: Computing Diversity. Communications of the ACM, 49,3, ACM, pp. 99 – 104.

[7] Gilbert, J. E. 2008. U.S. Patent No. 11,053. Washington, D.C.: U.S. Patent and Trademark Office.

[8] Gilbert, J.E., Eugene, W., Arbuthnot, K., Hood, S., Grant, M.M., West, M.L. & Swanier, C. (2009) Culturally Relevant Game Design: A Case Study for Designing Interactive Algebra Lessons for Urban Youth. i-manager's Journal of Educational Technology, 5,3, pp. 54-60.

[9] Hall, S. 1990. Cultural Identities and Diaspora. In Jonathan Rutherford (Ed.), Identity: Community, Culture, Difference, London: Lawrence and Wishart, pp. 222-37

[10] Hughes, T.P. 1993. Networks of Power: Electrification in Western Society, 1880-1930. The Johns Hopkins University Press, Baltimore.

[11] Irvine, J. J., & York, D. E. 2001. Learning styles and culturally diverse students: A literature review. In J. A. Banks (Ed.), Handbook of research on multicultural education (pp. 484-497). San Francisco: Jossey-Bass.

[12] Kamath, C., and Parker, A. 2000. Mining Data for Gems of Information. Research Highlight, Science and Technology Review, September 2000, pages 20-22. UCRL-52000-00-9 Retrieved from https://www.llnl.gov/str/Kamath.html

[13] Lee, C. D. 2007. Culture, Literature, and Learning: Taking Bloom in the Midst of the Whirlwind. New York, New York: Teachers College Press.

[14] Miller F.A. 1998. Strategic Culture Change: The Door to Achieving High Performance and Inclusion. Public Personnel Management. Vol. 27, No.2

[15] Swidler, A. 1986. Culture in action: Symbols and strategies. American Sociological Review, 51, 273-86.

[16] Young, P. A. 2008. The Culture Based Model: Constructing a Model of Culture. Educational Technology & Society, 11 (2), 107–118

Making an IMPACT on the environment: Sustainability Science and the I-School Movement

Fred Fonseca
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA, U.S.A.
01-814-865-6460
fredfonseca@ist.psu.edu

James Martin, Emeritus
Department of Psychology
The Pennsylvania State University
University Park, PA, U.S.A.
01-814-865-6460
jmartin501@gmail.com

Clodoveu Davis
Universidade Federal de Minas Gerais
Av. Presidente Antônio Carlos, 6627 – Belo Horizonte – MG – Brazil
clodoveu@dcc.ufmg.br

Gilberto Camara
Instituto Nacional de Pesquisas Espaciais
Av. dos Astronautas, 1758 – São José dos Campos – SP – Brazil
Gilberto.camara@inpe.br

ABSTRACT
Understanding how the environment is changing, in a global scale is one of the most important research questions of today. The sheer variety of areas of knowledge required to tackle this question is so great that only a solid interdisciplinary approach can succeed. Sustainability science aims at doing so staying at the intersection of more traditional research areas. The idea behind sustainability science is to develop ways to understand, integrate, and model the interaction between nature and society. The I-School movement is important for that purpose, considering its nature as a source of integration between diverse disciplines and research areas. Focusing mainly on modeling the interactions between nature and society, we opted to use a philosophical point of view to understand the implication s of putting together in a single model society and nature. We used Kant's view of man as phenomena (belonging to nature, being completely causally determined) and as noumena (human being as being free, as a thing in itself) to frame our discussion on how to build models that include both views. We also discuss the problem of integrating opposing views, such as society and nature, in a model, the Tower of Babel problem. We also discuss a common solution to this problem, the Newspeak solution, which is achieved through the imposition of a common ontology to which users are required to conform if they wish to participate at all. Looking for an integration of society and nature in modeling, we tie Gadamer's notion of Play to self-organization as a way to balance, within a single model, two contrary positions. Finally we conclude that a dialogue of clashing views can be held together without devolving into chaos, in which a contradiction implies all propositions, usually thought to be the consequence of bringing together inconsistent positions. This solution points beyond the either/or that is central to the Tower of Babel/Newspeak dilemma. The I-School movement has a unique opportunity to be the place where these discussions occur.

Categories and Subject Descriptors
10. [**Computer applications**]: II Physical sciences and engineering: Earth and atmospheric sciences

General Terms
Management, Design, Theory.

Keywords
Sustainability, multidisciplinarity, environmental modeling, philosophy.

1. INTRODUCTION
Global change has been the focus of much debate recently, due to clearly perceivable modifications of Earth's environment and climate. Divergent opinions and controversial research results, along with all the hype usually found in press coverage of this subject, indicate that scientists need to develop a better understanding of the complexity of physical-ecological-anthropogenic systems, developing a perception that the environment is influenced by a multitude of dynamic factors, originated form the interaction of natural and social systems.

Understanding how the environment is changing, in a global scale, is then one of the most important research questions of today. The sheer variety of areas of knowledge required to tackle this question is so great that only a solid interdisciplinary approach can succeed. Newly created fields, such as sustainability science [1, 2], have been gaining space precisely at the intersection of more traditional research areas. The idea behind sustainability science is to develop ways to understand, integrate, and model the interaction between nature and society.

The I-School movement is strategically posed to make a difference in sustainability research, because its multidisciplinary setting can support the understanding, representing and modeling global change, thus supporting the creation, application, and assessment of public policies for the environment. This paper presents a philosophical approach to the understanding of the interactions nature-society. Our belief is that the I-School movement has a unique opportunity to integrate the many disciplines necessary to address this challenge.

We opted to use a philosophical point of view to understand the implication s of putting together in a single model society and

nature. In section 2, we start by using Kant's notion that one can view humans as phenomena, objects of the natural sciences, and as noumena, things in themselves, not to be considered as a part of the causally integrated natural order of Nature. In section 3, we discuss the problem of integrating opposing views, such as society and nature, in a model. This problem was called elsewhere [3] the Tower of Babel problem. We also discuss a common solution to the Tower of Babel problem which is achieved through the imposition of a common ontology to which users are required to conform if they wish to participate at all. Fonseca and Martin [4] call this simplification the Newspeak solution. In section 4, looking for an integration of society and nature in modeling, we tie Gadamer's notion of Play to self-organization as a way to balance, within a single model, two contrary positions. Finally we conclude in section 5 that a dialogue of clashing views can be held together without devolving into chaos, in which a contradiction implies all propositions, usually thought to be the consequence of bringing together inconsistent positions. This solution points beyond the either/or that is central to the Tower of Babel/Newspeak dilemma.

2. PHILOSOPHICAL FOUNDATIONS FOR AN EPISTEMOLOGICAL PLURALISM IN MODELING SOCIETY AND NATURE

Among the many challenges for the I-School movement, one is how to integrate society and nature in the models of nature-society interactions. Next we take a look at some philosophical positions that can serve as a foundation for such models

In order to build scientific models of the interactions between society and nature we need to understand how humans are understood by science. Kant held that human beings could be seen from two complementary perspectives. According to the first perspective, one can view humans as phenomena, objects of the natural sciences, including those social sciences that adopt the methodologies and presuppositions of the sciences of nature. Especially important, here, is Kant's view that the sciences of nature take for granted the principle of causation – enunciated as every change of state is caused. This principle was central to Kant's categories of the understanding, which he believed to be given a priori and necessary to understanding the phenomena of Nature. Indeed, for Kant, Nature consists exclusively of phenomena that appear, either directly in perception, or indirectly, through the mediation of retroductive inference derived from directly given perceptions and previously established knowledge, all integrated and organized in terms of the a priori categories of the understanding.

According to the second perspective, Kant held that human beings, along with every other entity in nature, could also be considered as things in themselves, or noumena. In this case, an entity would not be considered as a part of the causally integrated natural order of Nature. This is possible to do because, according to Kant, the categories of understanding, including the principle of causation, are not derived from experience, but imposed, as presuppositions, upon the things experienced by the mind. Accordingly, it is possible to think of a human being as being free (as a thing in itself) without any self-contradiction, even though as an object of natural science, the same human being must be assumed to be completely causally determined.

In another context, some researchers [5] suggest that there is a hermeneutic connection between noumena and phenomena – agents as produced and as producers. But what are the larger implications of this connection? Fonseca and Martin [4] have suggested that it is possible to frame such fundamental hermeneutic oppositions in terms of the Gadamerian notion of play -- the mediating moment in Gadamerian hermeneutics. They argued that such play is the "place" where the clash between the "Tower of Babel problem" and what they have called its "Newspeak solution" might be addressed. Play allows for the full recognition of temporally distributed dialogue among clashing and mutually inconsistent perspectives, in contrast to such conditions of consistency as are usually associated with essentially atemporal consistent monologues. We suggest that the notion of Gadamerian play may be explicated in a way that brings together the Katian noumenal and phenomenal perspectives, thus giving a theoretical foundation to the creation of models that can held together the two perspectives.

3. THE TOWER OF BABEL PROBLEM AND THE NEWSPEAK SOLUTION

The Tower of Babel problem arises from the assumption that a necessary precondition of communication is the presupposition of a common logical or theoretical framework among those who would communicate. Making this assumption, the Tower of Babel problem might be solved by the imposition of a common ontology to which users are required to conform if they wish to participate at all. But such a maneuver would require considerable oversimplification of the world as it is represented on our models. Fonseca and Martin [4] call this simplification the Newspeak solution, after Orwell's Newspeak -- a revised version of English that was simpler and less capable of expressing different perspectives than traditional English.

The Tower of Babel problem has emerged as a fundamental barrier in the way of developing general and reusable models. The difficulty is that insofar as model designers attempt to accommodate, in the same system, groups of users possessing distinct ontological assumptions, they must address the problem of integrating information in ways that are compatible with the perspectives of all significant stakeholders. Of course, it might be possible to work out ad hoc solutions for a particular, limited set of ontological assumptions, but such a solution would be incompatible with the technological strategy, which aims at general and reusable models. A classic maneuver on the part of model designers is to adopt the Newspeak solution, i.e., when faced with the Tower of Babel problem, force all users to accommodate to a single perspective. In this case, the subtlety and ambiguity of differing perspectives is simply ignored. As in the case of Orwell's novel, implementation of the Newspeak

I-Conference'2010,

solution will likely require administrative authority to ensure that all users conform to the same ontological framework.

Both the Tower of Babel problem and the Newspeak solution share the assumption that communication requires a common underlying logical framework. We reject the relevance of either. Instead, we hold that communication takes place in a tacit and informal setting which is a necessary context, and ultimate source of all explicit, or potentially explicable, formal models. This context also makes negotiation across inconsistent perspectives possible. Such an informal context makes room for communication among persons who hold different perspectives.

A major weakness of the epistemic positions underlying the opposition between the Tower of Babel Problem and its Newspeak Solution is that neither of these positions has any well worked out account of the role of the tacit dimension in knowing. We have seen that both are grounded in what Karl Popper called the "myth of the framework" -- the assumption that communication depends on agreement concerning a common, explicable and logical/theoretical framework. Consequently, there is no account of the process of development and coming to understand of alternative perspectives. If one remains at the level of the fully explicit, it is difficult to see how a difference in points of view might be resolved. If one insists on bringing together inconsistent perspectives or facts, then by a well-known logical consequence, everything follows and the distinction between truth and falsehood is undermined. If on the other hand, in order to save the distinction between truth and error, one refuses to bring together inconsistent perspectives or facts, then one is faced with the choice between recognizing the existence of irreducibly incommensurable domains (relativism), and the elimination, as false, of all domains of facts not consistent with, and thus derivable from, a particular preferred domain (a rather narrow objectivism).

Despite its apparent efficiency, the fundamental problem with the Newspeak solution is that it cannot be implemented in situations where different users are required by the traditions of their own historical contexts to invoke different ontological assumptions. For example, Smith [3] points to the difficulties of integrating accounting systems when different users are required by the historically distinct traditions of case law to utilize different accounting structures. Even the same vocabulary items may have different meanings in different historical contexts. In such cases, differences in user orientation cannot be arbitrarily dismissed. They result from differences of history, which continue to constrain the interpretation of problems and the standards for solutions. They cannot be eliminated by the administrative fiat. Instead, they constitute what Gadamer would call “effective historical consciousness” – a concrete recognition of the effective role of history in constituting horizons from which we view events. In this effective historical consciousness, we become aware that the object is what it is from a perspective that we have arrived at as a result of our own history. But this does not entail a mere relativism. Instead, Gadamer is clear that “it is the task of effective historical consciousness to bring to explicit awareness the historical affinity” between the object of inquiry and the inquirer [6].

Next, in trying to explicate the connection between these two perspectives, we draw inspiration from Kant's Third Critique, the Critique of Judgment. We orient our discussion around Kant's revolutionary notion of self-organizing systems (SOS) – a notion that Kant introduced in order to make sense of living (biological) systems. In so doing, we hope to provide a more systematic relation between Kant's two epistemic perspectives, and with the aid of Gadamerian insights, move toward a third epistemic perspective that gives rise to the notion of wholes in which observers participate as acting and knowing constituents -- both acting and acted upon, both knowing and known.

4. GADAMERIAN PLAY, SELF-ORGANIZING SYSTEMS AND MODELING

It is important to understand that the introduction of Gadamerian hermeneutics at this point entails fundamental reconfigurations of Kant’s notion of SOS. Kant viewed the notion of SOS as a heuristic convenience for the study of biological systems. In contrast, we view the connection between noumena and phenomena in terms of Gadamerian Play -- explicated as a suitably amended SOS. From the perspective of Gadamerian hermeneutics, the SOS that embodies both noumenal and phenomenal moments is viewed as having fundamental ontological, and not merely heuristic, status. In consequence of these considerations, we see the values that direct the design, use, and continuing development of models as not merely subjective, but as dimensions of the SOS (or play) within which models emerge.

To our knowledge, the first thinker to explicitly introduce the notion of self-organization was Kant in the Critique of Judgment. In that work, he uses the idea of self-organization to characterize biological systems. Kant raises the issue of self-organization when addressing the notion of purposes. He is concerned with the notion of a being that is a purpose of nature. As opposed to an axe, which has a purpose only when considered in relation to the humans who create or use it, something, for Kant, is a purpose of nature when it is what it is because of what it is, not because of something else. Or, “I would say, provisionally, that a thing exists as a natural purpose if it is both cause and effect of itself (although [of itself] in two different senses.” [7] Considered as an example of a purpose of nature, a tree, Kant points out that a tree not only produces similar offspring, but it produces itself in that it sustains and furthers its own life. A tree is a self-organizing system. Furthermore, in self-organizing systems, he notes a reciprocal dependence of part and whole. The leaves of a tree are its production and a part of the tree. At the same time, the leaves are necessary for the continued life of the tree as a whole, which may be seen as being caused by the leaves. He concludes with the following definition, “In such a product of nature each part not only exists by means of all the other parts but is also regarded as existing for the sake of the others and of the whole, that is, as an instrument.” Here, the sense of “for the sake of the others” is intended to include the notion of production of the others. “An organized product of nature is one in which everything is a purpose and reciprocally also a means.” [7 p.255].

Current work with the notion of self-organizing or self-producing systems has been explored by Mingers [8], who

discusses the developments introduced by Maturana and Varela (in biology and theoretical psychology), Spencer-Brown (in logic and mathematics), Luhmann (social systems and the law), and their relationships with the epistemological views of Bhaskar (critical realism). Mingers points out that since Winograd and Flores' classic work [9] there has been relatively little done relating information systems modeling with self-organization. Central to Mingers' description of self-organizing systems is the contrast between self-organization and more traditional approaches to self-reference (e.g., Theory of Types). Speaking of the traditional approaches, Mingers holds, "All these approaches are similar in treating self-reference and its paradoxes as something to be avoided. In contrast, autopoietic theory treats these phenomena as central and constitutive of real systems," [8 p.156].

Our introduction of Gadamerian play is precisely in the spirit of Mingers' comments. What Mingers is describing is a system of mutually required oppositions. On the other hand, the play between contrary positions might be essentially self-organizing inasmuch as the paradoxical mutual requiredness is such that each of the two perspectives actually produces one another – asserting what the other presupposes. This, we suggest, is a fundamental link between the philosophic hermeneutics that underlies modeling contrary positions and the theory of self-organizing systems. Gadamerian, hermeneutical play constitutes a self-organizing ontology. Gadamer's notion of consciousness as, one the one hand, historically determined and thus limited, and on other hand, as capable of critical reflection upon those limits and thus free of them makes room for acknowledgment of both the phenomenal and noumenal aspects of the self, respectively.

5. CONCLUSIONS

In this paper we proposed a conceptual view of the possibilities of modeling the interactions of nature and society. This representation is fundamental for sustainability research in a global scale. The integration of diverse disciplines as proposed by the I-School movement put it in a unique position to develop ways to understand, integrate, and model the interaction between nature and society. We opted to discuss the challenges for science presented by the duality of man as being part of nature and having free will at the same time. We used Kant's view of man as phenomena (belonging to nature, being completely causally determined) and as noumena (human being as being free, as a thing in itself) to frame our discussion on how to build models that include both views.

We think that the antagonism of society-nature perspectives can be clarified by using Gadamer notion of Play understood as a self-organizing system. This play applied to modeling opposing perspectives indicates the possibility that support for a given purpose may derive from an antagonistic (contradictory) one. The notion of purpose – "for the sake of" – is thus derived from a larger notion of the self-organizing whole. Clearly, some purposes will prosper while others fail. The ones that prosper are those that are supported by the self-organizing system of which they are a proper part.

Seeking a characterization of Gadamerian play suited to the linguistically oriented hermeneutical situation of modeling society-nature interactions, we have discussed Kant's notion of self-organizing systems and the possibility of integrating noumenal and phenomenal perspectives in the same model. We think that this kind of integration exemplifies the sort of to and fro movement that Gadamer had in mind when describing the fundamental ontology within which understanding takes place. The kind of integration might be a way in which a dialogue of clashing views can be held together without devolving into chaos, in which a contradiction implies all propositions, usually thought to be the consequence of bringing together inconsistent positions. Rather than chaos, the new models created in this way might go on in an interesting and coherent way. In this way, the debate points beyond the either/or that is central to the Tower of Babel/Newspeak dilemma

6. REFERENCES

[1] Clark, W. C., Sustainability science: A room of its own, Proceedings of the National Academy of Sciences, 104, 6, pp. 1737, 2007.

[2] Clark, W. C. and Dickson, N. M., Sustainability science: The emerging research program, in Proceedings of the National Academy of Sciences, vol. 100: National Acad Sciences, 2003, pp. 8059-8061.

[3] Smith, B., Ontology, in The blackwell guide to the philosophy of computing and information, Floridi, L., Ed. Malden, MA: Blackwell, 2003, pp. 155-166.

[4] Fonseca, F. and Martin, J., Play as the way out of the newspeak-tower of babel dilemma in data modeling, presented at The 26th International Conference on Information Systems, Las Vegas, 2005.

[5] Monod, E., For a kantian foundation of is research: Proposals for an epistemological pluralism, presented at The Eighth Americas Conference on Information Systems, Dallas, TX, 2002.

[6] Bernstein, R. J., Beyond objectivism and relativism: Science, hermeneutics, and praxis. Philadelphia: University of Pennsylvania Press, 1983.

[7] Kant, I. and Pluhar, W. S., Critique of judgment. Indianapolis, Ind.: Hackett Pub. Co., 1987.

[8] Mingers, J., Self-producing systems: Implications and applications of autopoiesis. New York: Plenum Press, 1995.

[9] Winograd, T. and Flores, F., Understanding computers and cognition: A new foundation for design. Norwood, N.J.: Ablex Pub. Corp., 1986.

Reclaiming public spaces: Issues of visibility in ICT training for persons with disabilities in Latin America

Michele Maureen Frix
Technology & Social Change Group
University of Washington, Seattle
2511 E Yesler Way, Unit A
Seattle, WA 98105
1 206 617 7281
michelemfrix@gmail.com

Joyojeet Pal
ATLAS Institute
University of Colorado at Boulder
TASCHA
University of Washington, Seattle
4311 11th Ave NE, Seattle WA 98105
1 510 501 8679
joyojeet@uw.edu

Philip John Neff
Jackson School of Intl. Studies
University of Washington, Seattle
4746 18th Ave NE
Seattle, WA 98105
1 206 331 0561
phil.neff@gmail.com

ABSTRACT

As technology becomes more and more ubiquitous in the workplace and within social interactions, interest in the role of technology for the disabled has increased—first, in accessible design of technology itself, and second in the role of technology to enable people with disabilities to more fully participate in society. While work regarding accessible technology and disability has been conducted, it has been quite limited to developed countries. In this paper, we examine the use and perceived value of computer training centers throughout Latin America in order to understand how these services fit within the larger issues around employability and socio-economic exclusion of people with disabilities in society. We draw from in-depth research conducted with computer center users as part of the Organization of American States' POETA program in Mexico, Guatemala, Ecuador and Venezuela, as well as contrast such technical services with Colombian libraries which have sought to increase accessible technical services as a means of including people with disabilities in the broader public sphere. Using interviews with program and library administrators and users with disabilities, we explore the impacts of such training courses on employability, socio-economic exclusion of people with disabilities, and visibility of persons with disabilities in society. We find that libraries are a key point of discussion of the provision of technical training for people with disabilities when considering the perception that libraries are neutral, "safe spaces" in society which could allow for creating greater institutional awareness regarding the rights of people with disabilities.

Keywords:

Disability, ICTs, libraries, Latin America, visibility, employability.

1. INTRODUCTION

The United Nations High Commission on Human Rights reports that over 600 million people, approximately 10 per cent of the global population, are living with some type of disability. Over two-thirds of people with disabilities live in developing nations. It is estimated that between 80-90 per cent of people with disabilities in Latin America and the Caribbean are unemployed or living outside of the workforce, with 82 per cent of the population living in poverty.

In recent years, various factors including disability-related legislation, international conventions, academic departments of disability studies, and the work of thousands of community organizing groups and activists have raised mainstream recognition of issues of disability in terms of access. This period has coincided with tremendous developments in electronic technology that have impacted issues of disability in two important ways. First, there has been much work on access to technologies that have played a role in increasing the ability of people with disabilities to participate both socially and economically. Second, there has been much work in accessibility in daily living situations to expand the social participation of people with disabilities. While accessibility refers to a range of physical, communication, and virtual environments, the increasing ubiquity of personal technology has raised awareness on issues of accessibility on personal computers, mobile phones, and the internet.

There has been varied interest in technical and social science issues around technology and disability within the information studies realm on topics such as web accessibility [1], services in libraries [2], information organization [3] interfaces on accessible devices [4, 5], meta-data [6] and community informatics [7]. While the work in this space is diverse in areas of interest, the same is not true for geography. Both the work within information studies and more broadly in accessibility draws on empirical cases from a few industrialized nations. Work on disability in the developing world either comes from a global health burden perspective [8], or from the perspective of the social construction of disability [9]. Work on the role of accessible technology is virtually absent in studies of disability issues in the developing world.

In recent years, both disability rights groups and organizations working closely with them have recognized the importance of basic technology training for access to formal employment. In Latin America, a number of projects have been created in the past decade to provide subsidized technology training for people excluded from access to home computers [10]. Some such projects offer technology training and access to people with disabilities as a means of creating more equitable opportunities for them in the workplace. In this paper, we draw from extensive field research in Mexico, Guatemala, Ecuador, Venezuela, and Colombia to examine the use and perceived value of training at such centers, and to understand how these services fit within the larger issues around employability and social exclusion of people with disabilities.

We examine the work of several technology centers under two agencies in the region – the first, POETA[1] is a large multi-country project providing computer training for people with disabilities in 18 countries throughout Latin America and the Caribbean. Our research throughout all countries except Colombia is mostly at POETA-affiliated centers. In Colombia, we examine the work of the Bogota libraries.

2. METHODOLOGY

We use a qualitative interview-based methodology, with open-ended semi-structured interviews conducted over four months of field work. Our choice of using visibility as a central construct in our analysis is based on our preparatory field work with individual respondents and organizations in the course of our fieldwork. Issues of social and economic exclusion around disability are related to prevailing cultural issues, and in this environment both rights-based groups working on policy issues, as well as people with disabilities themselves, felt that their visibility in public space was a key first step towards greater inclusion.

Utilizing evidence from fieldwork, we seek to demonstrate our preliminary finding of the disconnect between the goals of individuals and their self-perception of employability in relation to ICT training, and the broader mission of social movements led by people with disabilities to emphasize the importance of visibility in the public sphere. The transcribed data from the field work is 1400 pages long. All interviews were conducted in Spanish.

2.1 Instrument Design

We conducted semi-structured, qualitative interviews, held primarily at the technology centers, offices, homes, libraries or in public places. The interview instrument was first developed in Seattle and iterated on the ground for the first week of interviewing to maintain a relatively consistent set of core questions across sites. In Colombia, we also conducted semi-structured focus groups with library users and we followed a similar set of questions to explore, although questions varied. The entire interview process, including briefing the respondent, took between 45 and 120 minutes per interview.

2.2 Sampling and Recruitment

Our selection of POETA centers and the Bogota libraries was based on two main criteria. First, we were interested in the role of Corporate Social Responsibility (CSR) in funding services for people with disabilities, thus both groups are selected from within beneficiaries of the Microsoft Community Affairs program. In this paper, we do not discuss issues around CSR and services for people with disabilities, but they are important to mention in explaining why we chose to select these two specific organizations.

Our second reason for selecting these two groups in our analysis is the focus of Bogota libraries on working within the state and institutional set-up, as contrasted with the POETA approach of establishing new technology centers for training of people with disabilities in which few centers link with government institutions.

With regard to the sampling of respondents, we used a snowball sampling method in all field sites. In each site we started with 3-5 respondents recruited through the training center followed by the remaining respondents recruited through the networks of the first set of respondents. We conducted research with primarily people with motor, auditory, and visual disabilities. In addition, we also interviewed program administrators, policy makers, and activists working in this space. The sampling varies from 5-20 interviewees at each location, and in all we conducted 150 interviews. The coding and analysis of the transcripts was done utilizing Atlas Ti.

2.3 Field Sites

We visited seven ICT training centers in Mexico, Venezuela, and Ecuador supported by POETA. They partner with local grantees in each country for up to two years. POETA grantees researched in our study included both public and private vocational training centers, universities, rehabilitation facilities, and NGOs primarily serving people with disabilities. POETA provides hardware and software including JAWS and screen magnifiers for people with visual impairments as well as some adaptive hardware. Program administrators often adapt curriculum from Microsoft's Unlimited Potential Program, which is required to be taught at every POETA center. Their courses seek to teach users the very basic functions of Microsoft operating system (nearly solely Windows XP) and the Microsoft Office Suite of programs. Users were also introduced to the Internet and e-mail programs. Some POETA partners teach modules in self-esteem, job-seeking strategies, resumes workshops, and socially acceptable, formal behavior in an office environment.

In Bogota, Medellin and Villavicencio, Colombia we visited four public libraries and three vocational and ICT training centers serving people with disabilities. Both the Bogota and Medellin network of libraries are winners of the Bill and Melinda Gates Access to Learning Award, in 2002 and 2009 respectively, for offering free access to ICTs in some of the country's poorest neighborhoods. Computer training courses for users with visual and auditory impairments focus primarily on training users in various assistive technologies, including scanners, screen reading technology known as JAWS, screen magnifiers, OCR (Optical Character Recognition) programs, as well as Braille. Programs at the libraries serving the disabled also include "invisible theater" courses, open meeting spaces, and workshops educating users on how to pursue further education.

3. DISCUSSION

3.1 Visibility in the Public Sphere

The disability activism movement has moved away from philanthropy and charity-based discourses towards those based on promoting disability rights and access in several parts of the western world through much of the 20th century. This transition has been relatively slower in the developing world, but calls by the UN Human Rights Commissioner to move away from charity in conceptualizing disability [11] and the specific terminology of rights in the United Nations Convention of the Rights of Persons with Disabilities has set the tone for greater recognition of equity in access to framing the international discussion on disability[2]. As Harlan Hahn argues,

"Features of architectural design, job requirements and daily life that have a discriminatory impact on disabled citizens…support a

[1] POETA: Partnership and Opportunities for Employment through Technology in the Americas

[2] United Nations. (2006) Convention on the Rights of Persons with Disabilities and Optional Protocol. <http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>

hierarchy of dominance and subordination between non-disabled and disabled segments of the population that is fundamentally incompatible with legal principles of freedom and equality…" [12]

Access to public space is a freedom or right which many non-disabled citizens may take for granted, however it is highly valued among the disabled community due to structural barriers, such as inaccessible transportation, and profound stigma and discrimination towards people with disabilities in society. The following quote highlights one interviewee's opinion of the public visibility of people with disabilities in Venezuela.

"You see many more people with disabilities, in jobs, in the street. People who have come out of their houses, most likely. It seems like the population of people with disabilities has grown, but I think it's not that it has elevated, but that they are leaving their homes. People see that there is a change, they are looked upon a bit better, and you have to take advantage of that."

'Juan', Venezuela (quadriplegia)

His opinion is optimistic and can be contrasted with that of another interviewee who uses a wheelchair in Ecuador.

"Yes, even the building where I work, where we are right now is not that accessible. For example, there is an elevator but when you need to leave the building there are steps so I always have to ask for help when I need to go up or down the steps. There is no ramp. There are many places where I always have to ask someone for help, or where there are steps or the entrance is very narrow."

'Adán', Ecuador (paraplegia)

Within this visibility framework, we look at programs, particularly Colombian libraries, which have implemented ICT training in order to not only give access to services, but also foster community building and respond to the need for computer training by the community of the disabled itself.

At this moment, *Biblioteca El Tunal* is one of few libraries in Bogota, Colombia which is fully accessible to people with disabilities, particularly the visually impaired and the deaf or hard of hearing. However, the community of people with disabilities has called upon policy makers and library administrators to make all libraries in the Bogota system accessible. By reconfiguring pre-existing institutional services in order to reach people with disabilities, the Bogota library system represents a shift in not only public policy, but also overall society, to allow people with disabilities an opportunity to be visible in a public space—the public library.

Libraries may be typically considered neutral "safe spaces" where community members are welcome; however, many library services are not accessible for people with disabilities, particularly technology services for people with visual impairments.

"They shut the doors in our faces at companies and refuse us employment so we wanted to convene here in the library as the National Association for the Blind in order to inform those companies that we are useful, productive people and just because we have a disability, that doesn't mean we are invalid… in order to do this we have to continue training and studying and so we that's why we come here to the library in order to stay up to date on systems and informatics training with the use of a program called Jaws."

'Martin', Colombia (visual impairment)

Jaws, the leading screen reader software from the U.S. publisher Freedom Scientific, is one of few tools which enables people with visual impairments to utilize a computer. It is also a significant barrier for people with disabilities, particularly those living in Latin America, due to a cost between US$1,000 and US$2,000. Very few technology training centers where our team conducted research had purchased original Jaws licenses, while most had acquired a licensed copy from ONCE-FOAL (The Spanish National Organization for the Blind's Foundation for Latin America), utilized a download demo version which required users to reboot every 45 minutes, or downloaded a pirated copy of the software.

"What has made things much easier is that through coming to the computer center in the library, I was able to burn a copy of a portable, more compressed version of Jaws, onto a CD and then I can go to a cyber café if I need to and I can work independently. But to be honest, I do not go to the cyber café though because I have the ease of being able to come to the library and use the services here…I practically live here."

'Claudia', Colombia (visual impairment)

Various beneficiaries of the computer training center in the Bogota Library referenced the sense of community they gained through attending courses and using the technical services offered. Despite the fact that some computer center users had access to a computer with Jaws at home, they still felt that coming to the library provided them with a community that was important not only to the advancement of their technical skills through computer training courses, but also to the rights-based movement led by people with disabilities. Visually impaired users travel from all around the urban area of Bogota to use the accessible technical services offered at the library. However, the library has come to represent much more than simply a public access point, but a place for people of all disabilities to convene, unify as a movement, and leverage the communal strengthen to reclaim human rights which were historically not granted to people with disabilities.

"Unfortunately, here in Bogota and Colombia in general, the disabled community is invisible in public policy and government priorities…people [with disabilities] stay invisible. They don't go out on protest. They stay silent…but alone as a community of the blind, we cannot do it…so what the library does for us is give us the ability to unify as a movement and it opens up opportunities and services to us. It is a great advancement to have access to services here. To be able to use the computer, the internet, all those services for free. It is a success that those here at Biblioteca el Tunal, weren't conformists."

'Micaela', Colombia (visual impairment)

3.2 Employability and Individual Capacity Building

The overwhelming statistics of unemployment and poverty among people with disabilities not only in Latin America, but throughout the world, has led to an increase in both public and private training centers which metric success based on employment rates of computer center graduates. This is also known as ones' *employability*, or "the ability to secure a job; the ability to keep an existing job or to improve that position in quality or income; the ability of beneficiaries to use elements of the training program as platforms to gain job experience if new to the labor market; and the ability to contribute to the overall productivity of business, government, and social labor." [13] We conducted research with several centers with similarly stated goals and carried out interviews with program beneficiaries, center administrators, and

potential employers. By contrasting such centers with the case study above of the Colombian libraries, we came to a preliminary conclusion that there is a disconnect between the goals of individuals and their self-perception of employability in relation to ICT training, and the broader mission of social movements led by people with disabilities to emphasize the importance of visibility in the public sphere. Typically, at computer training centers with a primary focus on ICT training and as a result, employability, individual capacity building is prioritized. The broader goals of including people with disabilities into the public sphere, building community, and the sharing of knowledge, become secondary goals. Many beneficiaries began attending the computer course in order to improve their individual skills and later seek employment. Basic computer competence was viewed by many users of centers focused on employability as an entry point into formal employment and a necessary tool in order to prove to employers and co-workers that a person with a disability was capable.

"I am not using Jaws there [at the post office] yet because I want to finish the course here first. Once I have finished the course here I will ask Agora to assist me in purchasing the license for Jaws so that I can install it on the computer at work…But I do not want to use it at my job yet because I want to learn a bit more, finish this course and I also want to prove to everyone at my job that I can do the job without it and once they trust me a bit more then I will install it. It is much easier for me to use the computer with Jaws and when my co-workers see me using it they are going to go crazy!"

'Felipe', Ecuador (visual impairment)

A theme among users interviewed in such centers was a self-perception of being more employable after acquiring ICT skills, but a feeling of disappointment once they attempted to actually enter the labor market. Few centers employed institutional champions, technical accompaniment services, or staff whose primary objective is to seek out potential employers, foster positive relationships with the community, and build a reputation of the center in order to bring in opportunities for program graduates.

"So we go, we explain to the company about our programs, we actually show them how they work because that's quite an important part because they always wonder how a blind person is actually able to use a computer because we are all accustomed to using a computer by sight."

'Jose', computer teacher for the visually impaired, Guatemala

However, such institutional champions were not commonly found. Many users had an expectation that once they had completed the course, they would be presented with opportunities of employment, but were faced with further discrimination.

"When I was doing the course, there were various companies for which one could submit their resume, and they would offer you employment. …That's what I was hoping for, to finish the course, and to enter a company a short time later, but that didn't happen."

'Andres', Venezuela (paraplegia)

"I also think that Mexicans are lacking a culture, not just Mexicans, but sometimes they see you have a disability and they think that you can't do things, or they believe you will give the business a bad image, or I don't know, that's what they imagine, that's what I have seen."

'Consuela', Mexico (motor impairment due to rheumatoid arthritis)

It was not only social stigma or structural barriers that prevented people with disabilities from finding work, but also national public policy towards people with disabilities, particularly with hiring quotas. In 2006, both Ecuador and Venezuela implemented new policies which are notable for mandating employers to hire people with disabilities, and for implementing inspections and substantial economic sanctions for non-compliance. Ecuador's reformed *Codigo de Trabajo* (Labor Code) establishes a percentage quota mandating the hiring of people with disabilities by all public and private employers with more than 25 employees, starting with 1 percent by the end of 2007 and reaching a maximum requirement of 4 percent by the end of 2010. The law is enforced by the Disabilities Unit of the Ecuadorean Ministry of Labor, and employers that fail to hire the required number of employees with disabilities are fined each month until they have met the requirement, at a rate equivalent to ten minimum salaries (the monthly minimum wage, set at $218 in 2009). However there were conflicting perceptions of such policies among users interviewed.

"What we need is a source of employment…there is a new law that says that every company must have at least three per cent of their workforce be people with disabilities but no one actually follows through with this. They are hiring people with disabilities in public companies but not in private companies because private businesspeople are not interested in collaborating with this kind of requirement or to support the handicapped population, or the disabled."

'Felipe', Ecuador (visual impairment)

4. CONCLUSION

The contrast between the Colombian libraries, which are public spaces and POETA's training centers, which are specialized spaces for people with disabilities, highlights an important area of contestation on issues of visibility. While both served very specific purposes, and it could easily be argued that the POETA centers were in themselves the best possible space for the services they provided, for respondents with disabilities themselves, the division between employability and visibility was in itself a spurious one. In a case of a rights-based approach where an awareness of social exclusion is an essential part of the movement for employability, the two are thus intricately tied and impossible to separate from one another.

It is, for instance, impossible to discuss employability in the context of Latin America without referring to the fact that most work places are not employing people with disabilities to perform the same functions as "normal" people. Thus, the use of public libraries in providing services for people with disabilities plays a larger and critical part in creating greater institutional awareness that necessarily comes with increasing employability.

Often, it was found that the employability metric was a goal imposed onto the institution by an outside funder. Previous work on ICTD and participatory design of computer training centers argues that in such development initiatives "the 'insiders' learn what the 'outsiders' want to hear…the needs become socially constructed and the dominant interests becomes community interests." [14] It is not to be said that many of such training centers also represented places of community, just as the library; the distinguishing factor is rather the primary purpose of creating the center in the first place. On one hand, the Bogota libraries system saw a warranted need by the disabled community to incorporate assistive technology into the libraries. A perception in society already exists that libraries are neutral, "safe spaces", open

to all members of the community, which makes it an easier transition for the non-disabled library users to acknowledge. In any discussion of the provision of training and technology access for people with disabilities, public libraries have to be a critical part of the discussion.

5. REFERENCES

[1] Paciello, M., *Web accessibility for people with disabilities.* 2000: Cmp.

[2] Schmetzke, A., *Web accessibility at university libraries and library schools.* Library Hi Tech, 2001. 19(1): p. 35-49.

[3] Kientz, J., et al. *Where's my stuff?: design and evaluation of a mobile system for locating lost items for the visually impaired.* 2006: ACM.

[4] Bigham, J., et al. *WebinSitu: a comparative analysis of blind and sighted browsing behavior*. 2007: ACM.

[5] Wobbrock, J., et al., *Text entry from power wheelchairs: EdWrite for joysticks and touchpads.* ACM SIGACCESS Accessibility and Computing, 2003: p. 110-117.

[6] Nikolova, S., et al. *Better vocabularies for assistive communication aids: connecting terms using semantic networks and untrained annotators.* 2009: ACM.

[7] Lazar, J., *Integrating accessibility into the information systems curriculum.* Proceedings of the international association for computer information systems, 2002: p. 373-379.

[8] Murray, C. and A. Lopez, *Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study.* The Lancet, 1997. 349(9064): p. 1498-1504.

[9] Yeo, R. and K. Moore, *Including disabled people in poverty reduction work:"Nothing about us, without us"*. World Development, 2003. 31(3): p. 571-590.

[10] Proenza, F., R. Bastidas-Buch, and G. Montero, *Telecenters for socioeconomic and rural development in Latin America and the Caribbean.* Washington, DC, Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO)/Banco Interamericano de Desarrollo (BID), mayo, 2001.

[11] Quinn, G., T. Degener, and A. Bruce, *Human Rights and Disability: The current use and future potential of United Nations human rights instruments in the context of disability.* 2002: United Nations Publications.

[12] Hahn, H., *The political implications of disability definitions and data.* Journal of Disability Policy Studies, 1993. 4(2): p. 41.

[13] Garrido, M., J. Sullivan, A. Gordon and C. Coward, *Researching the links between ICT skills and employability: An analytical framework.* CIS Working Paper No. 4, University of Washington, 2009.

[14] Bailur, S. *The complexities of community participation in ICT for development projects: The case of" Our Voices.* 2007.

Inviting Success:
Lessons from public access computing experiences around the world

Ricardo Gomez
The Information School
University of Washington
Mary Gates Hall 370, Box 352840
+1 206 685 1372
rgomez@uw.edu

word count: 6,600 (excluding references)

ABSTRACT
This paper presents findings from a comparative study of libraries, telecentres, and cybercafés in 25 countries around the world (and is part of a larger study in Latin America, Africa & the Middle East, South & Southeast Asia, and Eastern Europe); it focuses particularly on the factors that contribute to the centres' success across countries and types of centres. We clustered the results into five key success factors for public access computing: (1) understand and take care of local needs first,(2) build alliances with other venues, (3) collaborate with other media and community services, (4) strengthen sustainability, (5) train infomediaries and users. Taken individually, these factors are not new, as evidenced in the literature in the field. The value of these findings is their presentation together as a result of comparative research across multiple countries and different types of public access centres. This study provides strong validation that these five success factors are critical variables to be considered in policy decisions and program implementation. They also provide valuable direction for future research to explore each of the issues in more detail.

General Terms
Human factors, Performance

Keywords
telecentres, libraries, cybercafés, public access computing, success factors, information communication technologies development (ICTD)

1. INTRODUCTION
There is no magic formula for the success of a telecentre, a public library, or a cybercafé. They are places where people go use a computer, access the Internet, look for information, communicate with friends, and play games. These centres all contribute to wider access and the use of information and communication technologies (ICT) by underserved communities around the world. Each type of venue is different, and the context and experience of each of the 25 countries studied in this paper is different, too. Understanding that equitable access to, and meaningful use of, ICT plays an important role in social and economic development, especially in underserved communities (Warschauer 2003; Unwin 2009; Castells 2007), this paper seeks to answer the question: what are the key factors that contribute to the success of venues that offer public access to computers and the Internet, especially in underserved communities? Drawing from a large, international study, this paper offers a broad perspective

The last decade has seen an exponential growth of initiatives that offer public access to ICT as part of libraries, government and community centers, schools, cafés, and other small businesses. Most of the existing literature about public access ICT focuses on case studies and evaluations of telecentres and, to a lesser degree, public libraries and cybercafés. Tracking trends and drawing common lessons across countries and different types of centers is limited by the narrow focus of most studies, and by the wildly different research approaches and methods employed; furthermore, there is a plethora of success stories and anecdotes that illustrate specific examples of individuals, groups, and organizations transformed by newly gained access to ICT, but little systematic evidence of impact (Sey and Fellows 2009; Toyama, Reddy, and Saxenian 2006; Chinn and Fairlie 2007).

Through an international study in 25 countries, the Technology & Social Change group[1] of the University of Washington's Information School gathered detailed information about the current status, challenges, and lessons of public access computing across a broad spectrum of developing countries and emerging economies. Conducted by local research teams in each country, the study used a common research design and rationale to examine how and why people use public access venues, with a particular emphasis on the information needs of underserved and marginalized populations. This project is unique in that it covers a wide variety of developing countries around the world and looks at different types of public access venues using the same research design. Although studies of telecentres in specific development contexts have been conducted, there are few studies on public access to ICT in libraries, and almost none on cybercafés; no study has been done across different types of venues, and never across this many countries (see literature review below).

iConference 2010, Feb 3-6, 2010, Urbana-Champaign, Illinois, USA.

[1] Formerly known as Center for Information and Society, CIS.

The success of public access venues is defined differently in each context. In this paper, we provide a detailed analysis of the success factors and recommendations that emerged from each of the 25 countries, grouped into five common themes that are the most recurring across all types of venues and across all countries. Taken individually, these success factors are not new, as discussed in detail in this paper. But taken together, and as a result of original research across multiple countries and different types of public access centres, they provide valuable guidance for policymaking and program implementation, as well as valuable direction for future research to further explore each of the issues in more detail.

2. LITERATURE REVIEW

Even though it is clear that access to ICT alone does not automatically result in human development, it does enable new opportunities for bridging the digital divide[2] . Public access to ICT has become an increasingly important tool to promote more widespread access and use of ICT in developing contexts, as evidenced in both academic and industry literature on ICT and development (Heuertz et al. 2003; Kamssu, Siekpe, and Ellzy 2004; Selwyn 2003; van Dijk 2005; Bertot, McClure, and Jaeger 2008; Kuriyan and Toyama 2007; Toyama, Reddy, and Saxenian 2006; Wilson 2004).

While there have been many previous studies about public libraries, and ICT especially, in the US (Bertot et al. 2007; Walkinshaw 2007; Rutkauskiene 2008) and about telecentres for community development (Etta and Parvyn-Wamahiu 2003; Best and Kumar 2008; Kuriyan and Toyama 2007; Colle 2000; Proenza, Bastidas-Buch, and Montero 2002; Gomez and Hunt 1999) and, to a lesser degree, about cybercafés and their contribution to social and digital inclusion (Gurol and Sevindik 2007; Haseloff 2005; Robinson 2004; Rangaswamy 2008), we found no previous studies that have done systematic comparison of different types of venues and across multiple countries, or studies that extract common factors that enable success in public access venues from a broad variety of settings, as undertaken in this study.

By public access, we do not mean access to public or government information, but that the public has access to information and technology resources, irrespective of their geographic location, age, socio-economic status, education, gender, religion, nationality, culture, or race. Furthermore, public access does not preclude access in privately owned and operated locations, or places that charge a fee for use, as is the case in most cybercafés.

While the use of information and communication technologies (ICTs) are central to information access, the issue of ICT access as a solution proves to be trickier than one would imagine. DiMaggio and Hargittai (2001) indicate that:

> *In earlier work, the term "access" was used literally to refer to whether a person had the means to connect to the Internet if she or he so chose. More recently, "access" is sometimes used as a synonym for use. This is unfortunate, because studies that have measured both access and the extent of Internet use have found, first, that more people have access than use it; and, second, that whereas resources drive access, demand drives intensity of use among people who have access. (p. 2)*

Consequently, the Technology & Social Change group study perceives ICT access in a broader sense. Convergent with Warschauer (2003), we regard access not in the narrow sense of having a computer on the premises, but rather access in the much wider sense of being able to use ICT for personally or socially meaningful ends. I'm not sure any more, and rather than trying to go back to the source to figure it out I will delete it. In recent years, two concepts have been used with regard to ICT public access: universality and usability (Vanderheiden 2000; van Dijk 2006). Universality means that all human beings are entitled to access information, and usability is the potential of a device or service to be utilized to meet users' needs. However, universal access is still an aspirational goal, not a reality in most parts of the world.

Threats to equitable ICT access also prevent equitable social and economic development. "The unequal access to technology between groups due to differences in demography, economic status, and locations has been suggested to affect worldwide globalization through Internet connectivity" (Kamssu, Siekpe, and Ellzy 2004). The findings of our study uncover distinct patterns underlying the global disparities that ICT access carries. These disparities increase in developing countries. As van Dijk (2006) observed,

> *Development is uneven as well, and increasingly so, because the overwhelming majority of the population does not participate at all. It is lagging behind compared with the diffusion of new media in the nodes of their own countries, and even more as compared with the developed countries. This majority has little access even to old media such as the telephone, radio, TV and the press and to essential services such as electricity… The few computers and network connections in developing countries are barely used for applications in agriculture, health, education, public works, water resources, public transportation, public information, population planning, rural and urban land development or public utilities. (p. 252)*

3. METHODOLOGY

In this section, we briefly describe the research methodology used to collect and analyze the data in this study[3]. In making the methodological choices for the global study, we took into account the need for a common structure and approach to data collection, in order to enhance the comparability of the results, as well as the need for flexibility to adapt the research process to the needs and possibilities of each specific context.

[2] Some authors prefer the term "digital inclusion", while others prefer "community informatics" or "ICT for development". For a discussion on these labels see Gurstein (2008).

[3] Note that the complexity of this study cannot be fully accounted for in this short description. For a detailed description of the research methodology see Gomez (2009).

3.1 Selection of countries

Of 237 possible countries and territories in the world, the final 25 countries (Algeria, Argentina, Bangladesh, Brazil, Colombia, Costa Rica, Dominican Republic, Ecuador, Egypt, Georgia, Honduras, Indonesia, Kazakhstan, Kyrgyzstan, Malaysia, Moldova, Mongolia, Namibia, Nepal, Peru, Philippines, South Africa, Sri Lanka, Turkey, Uganda) were selected based on a careful process that used four successive sets of criteria to focus on a sample of developing countries with a mid-size geography and population, and with existing public library systems. The criteria for country selection were based on size, population and other demographic data[4], degree of freedom of expression[5] and political unrest[6], a measure of "needs and readiness" criteria[7], regional representation, and quality of country research teams. For a more detailed description of the country selection process and rationale see Gomez (2009).

3.2 Research Framework

An iterative research design (Barzilai-Nahon, Gomez, and Ambikar 2009) was conducted in two phases. The emergent insights and discussions from Phase 1 guided and sharpened the focus of Phase 2. From the outset, we identified a framework – Real Access – developed in South Africa by Bridges.org[8]. We adapted and refined Real Access, calling the resulting framework the Access, Capacity and Environment (ACE) Framework, and structured it as a tool to understand the range of economic, political, educational, infrastructure, cultural, organizational, and other factors that affect the way people use ICT in public access venues. The three pillars of this framework are: **equitable access**: physical access, suitability, and affordability of the venue, technology access; **human capacity**: human capacity and training (users and staff), meeting local needs, social appropriation; and **enabling environment**: socio-cultural factors, political will and legal and regulatory framework, popular support.

3.3 Data Collection

Nineteen local research teams were chosen (with some researchers representing more than one country) following an international call for proposals. Lead researchers from each team were brought together twice, at the beginning and halfway through the research process, to discuss the purpose, methodology, and emerging findings of the study. Each team conducted local research in local languages, using document reviews[9], expert interviews[10], site visits[11] , user surveys[12], operator interviews[13], and, in some cases, additional data gathering activities[14]. Detailed country reports were prepared by each local research team using a data-collection template designed to help teams organize their local fieldwork in order to answer detailed questions about Access, Capacity and Environment issues in each type of venue studied. The use of a common research design and methodology helped make data more comparable, even though the specific ways in which data was collected varied from one country to another in order to make it more locally relevant.

3.4 Data Analysis

After careful reading of all reports, we did a detailed annotation of success factors as they were represented in the data. During a facilitated workshop and several group discussions, we analyzed, grouped, and categorized the different findings across countries

[4] Size (exclude largest and smallest), population (exclude countries with population less than 1 million, and exclude highest population (India, China)), per capita income (exclude countries with per capita income over $11,116), human development index (HDI below 0.5)

[5] Based on Freedom House index: http://www.freedomhouse.org.

[6] Based on U.S. Dept. of State travel advisories.

[7] **Needs criteria**: Income inequality based upon Gini index (2006) from United Nations Development Program; ICT usage: based upon CIA World Factbook (2007); ICT cost: based upon International Telecommunications Union's World Information Society Report (2006). **Readiness criteria**: Politics: based upon World Economic Forum Global Information Technology Report (2006), Transparency International (2007), World Bank Worldwide Governance Indicators (2006); Skills: based upon International Telecommunication Union opportunity skills index (2007); ICT infrastructure: based upon International Telecommunication Union opportunity network index (2007).

[8] Non-profit organization based in South Africa, www.bridges.org.

[9] Document reviews: identified and reviewed salient literature in the country, including existing statistical information about population, ICT penetration, public access venues, government policies, and previous studies relevant to the study. On average, 30 to 50 documents per country were reviewed.

[10] Expert Interviews: identified at least ten specialists in the areas of interest of the project and conducted in-depth interviews with them. Interview guides were prepared in each country depending on the local needs and context. On average, 10 to 15 interviews with experts were conducted per country.

[11] Site visits: identified, visited, and observed six or more venues of each type (library, telecentre, cybercafé, or other). Site visits were undertaken for a minimum of a half day, making sure to include both urban and non-urban sites (ideally three of each). In selecting sites, research teams identified typical case samples of each type of venue, including both urban and non-urban sites. On average, there were 20 visits per country, and about 500 sites visited in total.

[12] User Surveys: user information was collected via a shared survey instrument. Each country team was allowed to add questions that they felt were relevant to the local context to enrich the overall body of evidence. At each site, every second or third user exiting the venue was surveyed. Teams surveyed between 40-50 users at each venue. Total users surveyed: 720-1100 per country. Given limited time and resources, user surveys were not intended to provide statistically significant samples of the population or of the venues studied, but an exploratory indication of trends and patterns for comparison and further research

[13] Operator Interviews: identified at least one operator in each site visited and held a structured interview to provide a more in-depth understanding of the venue, users and environment. Total operators interviewed: 18-22 per country.

[14] Additional optional data gathering: focus groups with users, operators or experts, additional visits and interviews, peer consultation and review.

and venues, all of which led to the formulation of the key success factors described in this paper. After finalizing the grouping and description of these factors, we went back to the detailed country reports to re-validate and document each one with examples from different countries. This process allowed us to combine multiple visions and readings of the rich data collected in the study, and resulted in higher-level, distilled lessons and success factors grounded in the data and the context of each country and venue.

Finally, we did a detailed re-reading and discussion of the country reports to identify and group trends in the data, selecting examples that best illustrated the key trends, and to make sure we did not miss any significant insights from local research partners.

3.5 Limitations of this Study

This study is groundbreaking in its breadth and scope; no other studies have systematically looked at different types of public access venues across multiple countries. Nonetheless, the breadth of the study also means that it does not provide an in-depth analysis of a particular venue, country, or experience, and findings cannot be easily generalized without a clear understanding of the specific context and the analytic framework used.

While the flexibility to translate and adapt the data collection tools to the needs and requirements of each country makes the study more locally appropriate, variations in the way data was collected or presented also makes the comparison of results across countries more problematic. The details discussed here may not be an exact reflection of any single country, but combined across all 25 countries the results represent a meaningful source of trends and patterns about success factors for public access ICT venues.

4. FINDINGS AND DISCUSSION

The following five themes were identified as the most salient and common factors that enhance the success of public access venues, with a particular focus on meeting the needs of underserved communities:

1. Understand and take care of local needs first
2. Build alliances with other venues
3. Collaborate with other media and community services
4. Strengthen sustainability
5. Train infomediaries and users

Each one is described and illustrated below, with examples from the study of libraries, telecentres and cybercafés in all 25 countries.

4.1 Understand and take care of local needs first

All 25 country reports noted the need for attention to the specific needs of underserved and rural populations. Successful implementation and maintenance of public access computing initiatives require a solid understanding of the information needs and resources of the communities they intend to serve. Most successful initiatives typically offer concrete solutions for specific issues of local contexts (i.e. their information needs) and the ability to build on existing practices in these communities. Community-needs assessment and social-development orientation are especially important if the public access initiatives are intended to reach underserved communities.

As stated by Schneiderman (2002), many people cannot benefit from technology "because of high cost, unnecessary complexity, and lack of relevance to their needs." This assessment is convergent with recent literature in the development field, where the concepts of participation, empowerment, and social capital are now fully integrated into development work (Cooke and Kothari 2001; Servaes and Malikhao 2005; Cadiz 2005). Meeting local needs is also a cornerstone of community approaches in the field of library and information science (Long 2001; Cooper 1993; Hillenbrand 2005; Aabo 2005; Worcester and Westbrook 2004), as well as the field of information and communication technologies for development (ICT4D) (Heeks 2009; Raiti 2007; Unwin 2009; Gurstein 2000).

The critical importance of understanding and serving local needs first is clearly reflected in the findings and recommendations of the researchers in the majority of the countries we studied. They show that for successful implementation of public access venues that serve local development it is important to have accurate data about the user community, their information needs, and the information systems already in use, as described in the following examples. Nearly all 25 country results echoed the results from the research teams in Algeria, Ecuador, and Namibia, who reported that while government efforts to expand ICT services are commendable, these efforts do not succeed if the ICT services fail to meet the needs of the local community (Bakelli 2008; Bossio and Sotomayor 2008; James and Louw 2008). Furthermore, researchers in former Soviet Republic countries, such as Georgia and Kyrgyzstan, noted that the extensive-yet-decrepit public library system in these countries no longer serves the community's actual needs (Ariunaa 2008; IPM 2008).

Knowing the current distribution of information systems and practices in a community is an important consideration as well. The researchers in Honduras, for instance, stressed the need to avoid duplication of efforts (Arias and Camacho Jiménez 2008). Malaysia's research team reported success in the distribution of venues throughout the country, including rural areas where the venues are incorporated into post offices, libraries, or health clinics, i.e. "places where local communities can access them easily" (Kushchu 2008).

In response to the need for specific requirements that will address local needs, suggestions from many of the country reports included building websites with local content information (health, environment, and agriculture) and websites for youth (focusing on education and knowledge building). Many countries, including Bangladesh, Sri Lanka, Peru, Namibia, South Africa, and Kazakhstan, deal with multiple languages spoken throughout the country. Georgia's research team, for instance, noted that information portals should disseminate information in both Armenian and Azerbaijani (IPM 2008). For certain regions in Peru, the team recommended online information should be more readily available in Quechua, an indigenous language spoken by a large proportion of the population (Bossio and Sotomayor 2008).

In addition to local-needs assessments, the local community needs to take ownership of the development of ICT programs and content, using them into create practical solutions that improve the lives of the individuals in the community. This idea of social appropriation of ICT services stands out across the 25 countries. Researchers in Sri Lanka recommended community involvement in order to "give ownership to the project and prevent it from

being a purely top down exercise" (Wanasundera 2008). In Argentina and the Dominican Republic, research teams reported that the population wanted to incorporate ICT use into their local reality and to leverage building community partnerships (Rozengardt and Finquelievich 2008; Alfaro, Molina, and Camacho Jiménez 2008). In Honduras, for example, community input has shown the potential to transform telecentres and libraries into spaces for knowledge exchange – meeting places not only for literacy training but also for "discussion, action and struggles" (Arias and Camacho Jiménez 2008).

Several research teams pointed out that ICT success could come from tapping the energy and skills of unemployed youth, who are more likely than others to use ICT services at cybercafés or community centers. In Uganda, the majority of ICT users were students and youth, with illiterate parents often relying on their children to fulfill their information needs (Ndaula 2008). Researchers in Sri Lanka, Namibia, and South Africa also recommended capitalizing on this untapped resource (James 2008; James and Louw 2008; Wanasundera 2008). By instructing unemployed youth in web-building and database maintenance, the local council could, in turn, employ them to increase and strengthen ICT services in their local community.

Many of the research teams emphasized that for ICT to reach and effectively serve local communities, they need to promote a positive information culture that includes constructive attitudes to information sharing and public awareness of ICT services. Public libraries in particular, are undermined by perceptions that they service students only, produce old and outdated information, or simply are not "cool" to visit. Mongolia's team described traditional libraries as places "where study happened, intellect was developed and newspapers were read" (Pact Mongolia 2008). Since people traditionally consider public libraries as a place to go for reading and accessing print materials, this public awareness campaign could start with creating a new image of public libraries. To address current perceptions of libraries as places strictly for students, other adult groups need to be made aware of the library's information services. Library outreach activities should also align with patrons' cultural and entertainment practices. The Dominican Republic's team, for instance, suggested that library coordinators should develop "fun" activities, such as organizing chess tournaments to draw people into the library (Alfaro, Molina, and Camacho Jiménez 2008).

Creating a positive-awareness campaign and taking calculated risks might revive public libraries from the "current state of decay, lack of capacity, and tired mentality" observed by researchers in Mongolia (Pact Mongolia 2008) and elsewhere. Moreover, public access venues need to address people's perceptions of information. The former Soviet republics have an extensive network of public libraries, but in Georgia, researchers found that many people believed they could not find high-quality information at the library because the building was poorly maintained (no heat, no funding). In addition, a widespread belief exists that the information provided at public libraries is outdated and of low quality (IPM 2008).

While libraries in these countries need to get additional funding, they also need to launch public relations campaigns to improve their image. For example, the Kazakhstan team recommended that the library system study the tactics used by banks in that country, which have been successful in raising public awareness of their mission and services (Pact 2008). In the same way, Moldova's researchers suggested that local public authorities, such as the mayor or local councils, get involved in the publicity campaigns. The involvement of local authorities would also help local governments become aware of the needs of underserved populations (OPINIA 2008).

4.2 Build alliances with other public access venues

Most research teams indicated that collaboration among and between different public access venues is limited but can yield powerful results if collaboration is promoted and strengthened. Networks of libraries, telecentre associations, and collaboration between cybercafés, all enhanced by partnerships between these venues in any particular location, will make public access to ICT stronger and more effective at serving the needs of local populations.

The collaborative model is convergent with current trends in understanding the power of networks as a distinctive characteristic of the information society. "Actors are no longer independent…They are dependent on each other. In networks, actors make agreements and more or less freely engage in associations. They cooperate on the basis of complementary strengths and they become *inter*dependent" (van Dijk 2006).

Most research teams in our study noted that collaboration can take many forms and lead to a variety of social impacts. Although this trend was noted across all countries, it was especially prevalent in Latin American countries. The Peruvian success factors for ICT included collaboration among similar venues: the "rich practice of association and networking of special libraries… linked by a common theme: AIDS, agriculture, forestry… [they] may have different goals, but they share some common problems and may share learning" (Bossio and Sotomayor 2008).

In Costa Rica, some telecentres have partnered successfully with libraries. Telecentres organized within libraries benefit from an established infrastructure and the ability of librarians to teach ICT literacy; in turn, libraries that host telecentres can use the Internet to supplement out-of-date library resources and better serve their communities (Sanchez González and Camacho Jiménez 2008). Likewise, Brazilian researchers suggested some innovative solutions to creating new visions of public libraries – the creation of libraries in telecentres and vice versa. For example, a library in the state of Bahia bought computers with support from Identidade Digital, a program that supports telecentres (Voelcker 2008).

From a different angle, Nepal's research team reported the use of a public/private partnership model where private, urban cybercafés serve "as capacity building and supporting partners for [public] telecentres in rural areas" (SAP International 2008).

4.3 Collaborate with other media and community services

Public access venues tend to be more successful if they extend partnerships and collaboration beyond public access venues to include other community services and media important to the community. Most notably, these collaborations include successful partnerships with community radio stations, health clinics, community organizations and government offices, as well as creative uses of mobile phones in combination with public access

venues. This model is convergent with current research in other domains of public services, and not limited to information alone. "Public services are now often provided by a complex network of partnerships, contracts, and alliances between government agencies, nonprofit organizations, and businesses, rather than by hierarchical government bureaucracy" (Huang and Provan 2007).

The networking and collaboration theme includes creative networking of existing resources of all types: human, equipment, connectivity, and experience. Namibia's research team, for example, encouraged further utilization of established information kiosks in Community Information Resource Centers for different community development activities (James and Louw 2008). Researchers in South Africa noted the potential for leveraging an existing program: "The scoping of the HIV/AIDS centers strongly suggests that there is an opportunity to explore a programmatic intervention by ICT funders in partnership with one or more of the HIV/AIDS programs discussed" (James 2008).

Similarly, researchers from several of the 25 countries offered innovative, concrete suggestions of technological communication devices other than computers, most notably radio and mobile phones. As Kazakhstan's team explained, "Combining various media types allows maximizing the impact and ensuring all groups involved are covered. Radio may not be appealing to young Internet users while rural elderly population will never choose [a different] option" (Pact 2008). Teams studying the Philippines and Argentina both observed that these countries have greater access to cell phones than computers and, therefore, recommended expansion of government services through cell phones. The Philippines' research team, with its list of six "mobile phone applications" recommendations, specifically advocated for the expansion of text messaging services with development-oriented information (Ideacorp 2008; Rozengardt and Finquelievich 2008). Two other research teams proposed ideas for integrating these "other" existing technologies into communities: Uganda's researchers argued for "strategic establishment of a community radio at every Public Library (PL) facility per district" (Ndaula 2008); while Mongolia's team promoted "the range of information vectors (including radio, TV and mobile phone) that can be developed at community level" (Pact Mongolia 2008). Our study did not explore the interaction of ICT in public access venues with other technologies such as community radio or mobile phones; additional research would be warranted to get a better understanding of the opportunities presented by better collaboration with other media.

4.4 Strengthen Sustainability

Sustainability of public access venues is a critical issue that touches on multiple dimensions: financial, technical, social, and cultural. Government funding and support for public libraries has been declining in many countries, and donors' interest in telecentres has declined as well, threatening the financial sustainability of these public access venues. Successful telecentres have found creative ways to generate revenues, and popular libraries have found creative ways to build strong community support. But local community involvement alone cannot ensure the sustainability of public access ICT. Governments must also work to create an environment that strengthens and sustains public access to information and ICT resources if they are to meet the needs of underserved communities.

Challenges to sustainability have been extensively reported in the literature about public access to ICT, especially for the telecentres (Delgadillo, Gomez, and Stoll 2002; Bailey 2009; Best 2008; Gordon, Moore, and Gordon 2004; Gurstein 2005; Jensen and Esterhuysen 2001; Proenza 2001; Stoll and Menou 2003; Toyama et al. 2005). Many telecentre projects have simply failed after the original donors have left. Mayanja (2006) observed, "financial and social sustainability of telecentres remains one of the key challenges of the digital inclusion programming more than a decade after."

In an editorial of the *Journal of Community Informatics* dedicated to telecentre sustainability, Michael Gurstein (2005) suggests:

> *What is meant by "sustainability" in the ICT context is less a matter of a broad "configuration of civilization" and more to do with day to day slogging by community members in meeting the payroll and keeping the machines running amidst the wear and tear of daily life (both physical and electronic) while always keeping in mind how the technology could be used to respond to the needs (and opportunities) of their local communities. … When we are speaking of "sustainability" in the context of ICTs we should perhaps be speaking of "sustainabilities" rather than "sustainability", for there are many dimensions of this issue which go much beyond the simple economic and the meeting of weekly payrolls.*

As succinctly summed up by researchers in Costa Rica, the "digital divide is only a small part of the economic divide" (Sanchez González and Camacho Jiménez 2008). When governments plan and implement ICT services, they should be mindful of the needs of disenfranchised and marginalized communities. Kazakhstan's research team advocated "affirmative action" to serve the needs of marginalized groups in order to create a more inclusive information society (Pact 2008). The country's Program on Reduction of Information Inequity has so far failed to identify vulnerable groups, such as the homeless or the disabled. These groups in particular need extra assistance to access information, including finding government services.

In our study, most research teams pointed out the importance of having government departments devoted to ICT development. Collaborating with other governmental units – a "Ministry of ICT," as it is called in Colombia (Universidad de los Andes 2009) – could oversee the provision of online content regarding citizens' rights and governmental services. Argentina's research team argued for the adoption of a transparent e-government concept: "Public information venues could become privileged places of training citizens to participate in E-Government and E-Democracy processes" (Rozengardt and Finquelievich 2008). Namibia's team advocated for more venues where citizens could access government information free of charge, such as information kiosks at Community Information Resource Centers (James and Louw 2008).

In addition to financial and political sustainability, technological sustainability needs to be ensured by making technology work in low-resource environments. Public access venues aimed at underserved communities frequently face particular technical limitations of working in low-resource environments: poor electricity, connectivity, and outdated technology make it

especially hard to operate effectively. Making ICT sustainable anywhere obviously requires basic infrastructure: electricity, equipment, and Internet connections. This infrastructure also includes support systems − technical support, troubleshooting, networks − to maintain information systems and ensure that they function efficiently, even in environments where resources are scarce.

Many countries highlighted the need for electricity and basic infrastructure to support ICT. Researchers in Bangladesh credited the relative success of urban (as opposed to rural) ICT venues to the availability of an uninterrupted power supply. The reports from Algeria, Ecuador, Georgia, Kyrgyzstan, and Peru all called for increased support of basic infrastructure in rural areas (Ariunaa 2008; Bakelli 2008; Bossio and Sotomayor 2008, 2008; IPM 2008). Even where buildings, electricity, and computers were available, Internet access and bandwidth were problematic. The Bangladesh research team noted that "the performance of the venues with Internet connection is way better than the venues without Internet connection" (Development Research Network 2008). In Brazil, researchers identified infrastructure in the form of "updated equipment (adequate computers, Internet bandwidth)" as a key success factor for ICT (Voelcker 2008).

Beyond basic infrastructure, further analysis of the success factors and recommendations revealed that many of the research teams addressed additional issues of sustainability. The Costa Rican team attributed the failure of many rural telecentres to the challenges beyond installation, including maintenance of the equipment and software updates. Because only government technicians are permitted to repair equipment or address software problems and viruses at these telecentres, many of them have only two out of six computers working at any given time while they await technical support (Sanchez González and Camacho Jiménez 2008). The Bangladesh team expressed this problem as a need to "strengthen the support system (technical, know-how, and operational) for the public access venues" (Development Research Network 2008).

Maintenance is only part of the true cost of sustainable infrastructure. Ongoing costs must be considered in addition to initial investment. Researchers in Namibia declared that "the cost of computers and their software is limiting their availability. Government should therefore have a policy to support the use of Free and Open Source software (FOSS)" (James and Louw 2008). In Bangladesh, where the availability of electricity in rural areas is "dismal" and unlikely to change soon, the recommendation is for an investment in "low power consuming device[s] with higher battery life" in order to bring ICT services to the public (Development Research Network 2008). These recommendations point to the need for forethought and planning in order to make technology work in low-resource environments.

4.5 Train users and infomediaries

The fifth and last theme in the success factors that emerged from our study in 25 countries has to do with training users and operators of the public access venues. If communities are to benefit from public access to ICT, both users and operators need to have the basic training and know-how in order to use and operate the services. Building this capacity starts with basic literacy (reading and writing) training and includes basic digital literacy (use of computer, its basic applications and features). Strengthening the training and capacities of librarians and other operators of public access venues is also critical to the operator's success, especially if they are to provide guidance, training, and support services to users, directly or indirectly. Trained and motivated librarians and operators make better information brokers, or "infomediaries," who help make information resources more meaningful to the local communities, and help bring local knowledge and information resources to the public access venues.

Formal infomediaries include librarians and operators of telecentres and cybercafés. As part of a broader literature review on ICT impact, Sey & Fellows point out that infomediaries "have been found to be important contributors to the viability and sustainability of a public access venue" (Sey and Fellows 2009) and critical to the success of telecentres in particular (Bossio 2004; Best and Kumar 2008; Gomez and Hunt 1999; Parkinson 2005; Whyte 2000).

Extending the notion beyond the formal role of librarians or telecentre operators, other informal infomediaries play a critical role as well. Abrahamson & Fisher (2007) describe this informal role as lay information mediary behavior (LIMB), for example a person who finds information for another member of the family or for a friend or neighbor. Expanding out further, social networks also play a crtical role in information facilitation. Schilderman (2002) suggests that "social networks are the foremost source of information of the urban poor" (p. 4) and that the poor tend to believe people they trust rather than perhaps more informed contacts with which they do not have close ties. He then develops the concept of "key informants" (aka "infomediaries") defined as "people inside, or sometimes outside, a community who are knowledgeable in particular livelihoods aspects, and are willing to share that knowledge" (Schilderman 2002). In order to tap into this resource to help serve the information needs of this underserved population, he cited a number of success factors, including: involvement of the poor themselves as equal partners, building on local knowledge, the use of community-based communication methods, and building the capacity of community-based organizations and key individuals within them. He then identified seven key characteristics of effective key infomediaries (Schilderman 2002):

- their capacity to provide information in an accessible format
- their willingness to share information rather than hold onto it
- their ability to get hold of information and adapt it to a local context
- their experience, education, knowledge and reliability
- their accessibility, proximity and helpfulness
- their social sensitivity and capacity to involve residents
- their leadership qualities, influence and moral authority

Training of users and, more importantly, of infomediaries (both formal and informal) is a strong common success factor across all 25 countries in our study. Honduras researchers described ICT training as "elemental" to success. They also suggested that the success of cybercafés ought to be passed along to society by taking responsibility for training the population in the use of ICT, thereby "boosting the capacities of the individuals and generat[ing] a major communal impact" (Arias and Camacho Jiménez 2008). Researchers in Indonesia took the call for

increased digital literacy a step further, advocating universal free ICT training for all, especially for underserved populations, as critical to success (Kushchu 2008). The Argentina team pointed out that information literacy training for users should encompass their real interests and needs in order "to make a real appropriation of ICTs" (Rozengardt and Finquelievich 2008). Similarly, the Ecuador team called for the development of ICT training programs that address the needs of special groups, such as "women, illiterates, non-Spanish speakers and older people" (Bossio and Sotomayor 2008).

Researchers in Georgia extended the call for training to include venue operators, who should learn more about searching for health and education information (IPM 2008). The Malaysia research team listed centralized training for venue operators as one of its main success factors. Along these lines, researchers in Kyrgyzstan noted the need to "renew training and education curriculum of the ICT specialists to meet requirements of fast growing industry" (Ariunaa 2008). Another group who could benefit from training is local businesses; Indonesia's team recommended that the government should support local e-commerce by training "small to medium businesses to enable them to upload their products to the Warmasif [telecentre] website" (Kushchu 2008). The Moldova team argued that librarians and venue operators should be trained in both fundraising and grant proposal development in order to acquire more financial support for ICT programs (OPINIA 2008).

These different kinds of "infomediaries" − venue operators, librarians, government and community leaders, businesspeople, etc. − take part in a developing system of ICT knowledge training that would ideally extend throughout the whole population. Identifying the need to "train and deploy digital information facilitators to create and meet local information needs," Mongolia's team envisioned an investment in human resources that would benefit the whole country (Pact Mongolia 2008). Researchers in Kazakhstan reported a lack of human capital and a market demand for IT specialists (such as venue operators), which is five to seven times higher than current capacity due mostly to a lack of qualified teachers and quality education in schools and universities (Pact 2008).

5. CONCLUSIONS

We have presented five key factors that contribute to the success of venues that offer public access to ICT: understand and take care of local needs first, build alliances with other venues, collaborate with other media and community services, strengthen sustainability, and train infomediaries and users. These five factors are not new. What is new is to see them confirmed as they emerge from a broad empirical study of libraries, telecentres and cybercafés in 25 countries. This kind of validation constitutes a solid statement to policymakers and practitioners to help focus their efforts where they can make the most difference to the communities they intend to serve. Furthermore, our findings provide clear direction for future research on public access to ICT. Future research can help provide a better understanding of the local manifestations of each of the success factors we analyzed, and of the implications of these trends for measuring the impact of public access to ICT for underserved communities around the world.

6. REFERENCES

[1] Aabo, Svanhild. 2005. The role and value of public libraries in the age of digital technologies. *Journal of Librarianship and Information Science* 37 (4):205-211.

[2] Abrahamson, J., and K. E. Fisher. 2007. What's past is prologue: Towards a general model of lay information mediary behaviour. *Information Research* 12 (4).

[3] Alfaro, Francia, José Pablo Molina, and Kemly Camacho Jiménez. 2008. Public access to information & ICTs: Dominican Republic. In *Public Access Landscape study final report*. Seattle: presented by Sulá Batsú to University of Washington Center for Information & Society (CIS).

[4] Arias, Melissa, and Kemly Camacho Jiménez. 2008. Public access to information & ICTs: Honduras. In *Public Access Landscape Study final report*. Seattle: presented by Sulá Batsú to University of Washington Center for Information & Society (CIS).

[5] Ariunaa, Lkhagvasuren 2008. Public access to information & ICTs final report: Kyrgyzstan. In *Public Access Landscape Study final report*. Seattle: University of Washington Center for Information & Society (CIS).

[6] Bailey, Arlene. 2009. Issues affecting the social sustainability of telecentres in developing contexts: A field study of sixteen telecentres in Jamaica *The Electronic Journal on Information Systems in Developing Countries* 36 (4):1-18.

[7] Bakelli, Yahia. 2008. Public access to information & ICTs final report: Algeria. In *Public Access Landscape Study final report*. Seattle: presented to University of Washington Center for Information & Society (CIS).

[8] Barzilai-Nahon, Karine, Ricardo Gomez, and Rucha Ambikar. 2009. Conceptualizing a Contextual Measurement for Digital Divide/s: Using an Integrated Narrative. In *Overcoming Digital Divides: Constructing an Equitable and Competitive Information Society*, edited by E. Ferro, Y. Dwivendi, G. Ramon and M. Williams: Idea Group Inc.

[9] Bertot, John C., Charles R. McClure, Susan Thomas, Kristin M. Barton, and Jessica McGilvray. 2007. Public Libraries and the Internet 2007: Report to the American Library Association. Tallahassee, FL: College of Information, Florida State University.

[10] Bertot, John Carlo, Charles R McClure, and Paul T Jaeger. 2008. The impacts of free public Internet access on public library patrons and communities. *Library Quarterly* 78 (3):285-301.

[11] Best, Michael and Rajendra Kumar. 2008. Sustainability Failures of Rural Telecenters: Challenges

from the Sustainable Access in Rural India (SARI) Project. *Information Technologies & International Development* 4 (4):14.

[12] Best, Michael, and Rajendra Kumar. 2008. Sustainability Failures of Rural Telecenters: Challenges from the Sustainable Access in Rural India (SARI) Project. *Information Technologies & International Development* 4 (4):14.

[13] Bossio, Juan Fernando. 2004. Social Sustainability of Telecentres from the Viewpoint of Telecentre Operators: A Case Study from Sao Paulo, Brazil, Economics, London School of Economics, London.

[14] Bossio, Juan Fernando, and K. Sotomayor. 2008. Public access to information & ICTs final report: Ecuador. In *Public Access Landscape Study final report*. Seattle: presented by Alfa-Redi toUniversity of Washington Center for Information & Society (CIS).

[15] ———. 2008. Public access to information & ICTs final report: Peru. In *Public Access Landscape Study final report*. Seattle: presented by Alfa-Redi to University of Washington Center for Information & Society (CIS).

[16] Cadiz, C.M. . 2005. Communication for empowerment: The practice of participatory communication in development. In *Media and glocal change: Rethinking communication for development*, edited by O. Hemer and T. Tufte. Göteborg, Sweden: NORDICOM.

[17] Castells, M. 2007. Communication, power, and couter-power in the network society. *International Journal of Communication* 1:238-266.

[18] Chinn, D.M., and W.R. Fairlie. 2007. The determinants of the global digital divide: A cross-country analysis of computer and internet penetration. *Oxford Economic Papers* 59 (1):16-44.

[19] Colle, Royal D. . 2000. Communication Shops and Telecentres in Developing Countries. In *Community Informatics: Enabling Communities With Information*, edited by M. Gurstein. Hershey: Idea Group Inc.

[20] Cooke, B., and U. Kothari, eds. 2001. *Participation: The new tyranny?* London: Zed.

[21] Cooper, S. M. 1993. *Community analysis methods and evaluative options: The CAMEO handbook.* Richmond, VA: VA State Library and Archives.

[22] Delgadillo, Karin, Ricardo Gomez, and Klaus Stoll. 2002. Community telecentres for development : lessons from community telecentres in Latin America and the Caribbean. IDRC, Ottawa.

[23] Development Research Network. 2008. Public access to information & ICTs final report: Bangladesh. In *Public Access Landscape Study final report*. Seattle: presented to University of Washington Center for Information & Society (CIS).

[24] DiMaggio, P.J., and E. Hargittai. 2009. *From the digital divide to digital inequality: Studying internet use as penetration increases* 2001 [cited September 09 2009]. Available from http://www.princeton.edu/~artspol/workpap.html.

[25] Etta, Florence, and Sheila Parvyn-Wamahiu. 2003. *Information and communication technologies for development in Africa: volume 2. The Experience with Community Telecentres*. Ottawa/Dakar: International Development Research Centre (IDRC) /Council for the Development of Social Science Research in Africa.

[26] Gomez, Ricardo. 2009. Structure and Flexibility in Global Research Design: Methodological Choices in Landscape Study of Public Access in 25 Countries. In *CIS Working Paper no. 8*. Seattle: University of Washington.

[27] Gomez, Ricardo, and Patrik Hunt, eds. 1999. *Telecentre Evaluation: A Global Perspective*. Ottawa: IDRC.

[28] Gordon, Margaret T., Elizabeth J. Moore, and Andrew C. Gordon. 2004. Sustainability in First Ten States to Receive Gates Awards: Most Libraries Maintaining Public Access Computing Programs, but 25% Are Still Fragile. Seattle, WA: Evens School of Public Affairs of the University of Washington.

[29] Gurol, Mehmet, and Tuncay Sevindik. 2007. Profile of Internet Cafe users in Turkey. *Telematics and Informatics* 24 (1):59-68.

[30] Gurstein, M. . 2000. Community informatics: Enabling communities with information and communications technologies. In *Community informatics: Enabling communities with information and communications technologies* edited by M. Gurstein. Hershey, PA: Idea Group.

[31] ———. 2005. Editorial: Sustainability of community ICTs and its future. *The Journal of Community Informatics* 1 (2):2-3.

[32] ———. 2008. *What is community informatics (and why does it matter)?* Milan, Italy: Polimetrica.

[33] Haseloff, Anikar M. 2005. Cybercafes and their Potential as Community Development Tools in India. *The Journal of Community Informatics* 1 (3):13.

[34] Heeks, Richard. 2009. The ICT4D 2.0 Manifesto: Where next for ICTs and International Development? Manchester, UK: Institute for Development Policy and Management.

[35] Heuertz, L., A. C. Gordon, M. T. Gordon, and E. J. Moore. 2003. The impact of public access computing on rural and small town libraries. *Rural Libraries* 23 (1):51-79.

[36] Hillenbrand, C. 2005. Librarianship in the 21st century - crisis or transformation? *Australian Library Journal* 54:164-181.

[37] Huang, K., and G. K. Provan. 2007. Resource tangibility and patterns of interaction in a publicly funded health and human services networks. *Journal of Public Administration Research and Theory* 17 (3):435-454.

[38] Ideacorp. 2008. Public access to information & ICTs: Philippines. In *Public Access Landscape Study final report*. Seattle: presented to University of Washington Center for Information & Society (CIS).

[39] IPM. 2008. Public access to information & ICTs final report: Georgia. Seattle: University of Washington Center for Information & Society (CIS).

[40] James, Tina, et al. 2008. Public access to information & ICTs final report: South Africa. Seattle: University of Washington Center for Information & Society (CIS).

[41] James, Tina, and Milton Louw. 2008. Public access to information & ICTs final report: Namibia. Seattle: University of Washington Center for Information & Society (CIS).

[42] Jensen, Mike, and Anriette Esterhuysen. 2001. *The Telecentre Cookbook for Africa: Recipes for self-sustainability*. Paris: UNESCO.

[43] Kamssu, J.A., S.J. Siekpe, and A.J. Ellzy. 2004. Shortcomings to globalization: Using Internet technology and electronic commerce in developing countries. *The Journal of Developing Areas* 38 (1):151-169.

[44] Kuriyan, Renee, and Kentaro Toyama. 2007. Review of Research on Rural PC Kiosks. http://research.microsoft.com/research/tem/kiosks/.

[45] Kushchu, Ibrahim. 2008. Public access to information & ICTs final report: Indonesia. Seattle: University of Washington Center for Information & Society (CIS).

[46] ———. 2008. Public access to information & ICTs final report: Malaysia. Seattle: University of Washington Center for Information & Society (CIS).

[47] Long, S. A. 2001. Libraries Build Community. *Jiao yu zi liao yu tu shu guan xue = Journal of educational media & library sciences = EMLS*. 39:15-22.

[48] Mayanja, M. . 2006. Rethinking telecentre sustainability approaches How to implement A social enterprise approach: Lessons from India and Africa. *The Journal of Community Informatics* 2 (3).

[49] Ndaula, Sulah. 2008. Public access to information & ICTs final report: Uganda

[50] Seattle: University of Washington Center for Information & Society (CIS).

[51] OPINIA, Independent Sociological and Information Service. 2008. Public access to information & ICTs final report: Moldova. Seattle: University of Washington Center for Information & Society (CIS).

[52] Pact. 2008. Public access to information & ICTs final report: Kazakhstan. Seattle: University of Washington Center for Information & Society (CIS).

[53] Pact Mongolia. 2008. Public access to information & ICTs final report: Mongolia. Seattle: University of Washington Center for Information & Society (CIS).

[54] Parkinson, Sarah. 2005. Telecentres, Access and Development: Experience and Lessons from Uganda and South Africa. IDRC.

[55] Proenza, Francisco. 2001. Telecenter Sustainability - Myths and Opportunities. *Journal of Development Communication* 12 (2, Special Issue on Telecentres):15.

[56] Proenza, Francisco, Roberto Bastidas-Buch, and Guillermo Montero. 2002. Telecenters for Socioeconomic and Rural Development in Latin America and the Caribbean. Inter-American Development Bank. http://www.iadb.org/sds/itdev/telecenters/exsum.pdf.

[57] Raiti, Gerard C. 2007. The lost sheep of ICT4D research. *Information Technologies and International Development* 3 (4):1-7.

[58] Rangaswamy, Nimmi. 2008. Telecenters and Internet Cafes: the Case of ICTs in Small Businesses. *Asian Journal of Communication* 18 (4):22.

[59] Robinson, Scott. 2004. Cybercafés and national elites: constraints on community networking in Latin America. London: Community practice in the network society.

[60] Rozengardt, Adrián, and Susanna Finquelievich. 2008. Public access to information & ICTs final report: Argentina. Seattle: University of Washington Center for Information & Society (CIS).

[61] Rutkauskiene, Ugne. 2008. Impact measures for public access computing in public libraries. Vilnius University.

[62] Sanchez González, Adriana, and Kemly Camacho Jiménez. 2008. Public access to information & ICTs final report: Costa Rica. Seattle: University of Washington Center for Information & Society (CIS).

[63] SAP International. 2008. Public access to information & ICTs final report: Nepal. Seattle: University of Washington Center for Information & Society (CIS).

[64] Schilderman, Theo. 2002. Strengthening the knowledge and information systems of the urban poor. Dept. for International Development (DFID).

[65] Selwyn, N. 2003. ICT for all? Access and use of public ICT sites in the UK. *Information, Communication & Society* 6 (3):350-375.

[66] Servaes, J., and P. Malikhao. 2005. Participatory communication the new paradigm. In *Media and glocal change: Rethinking communication for development* edited by O. Hemer and T. Tufte. Göteborg, Sweden: NORDICOM.

[67] Sey, Araba, and Michelle Fellows. 2009. Literature Review on the Impact of Public Access to Information and Communication Technologies. In *Working Paper No. 6*. Seattle: Center for Infromation & Society, Univ. of Washington.

[68] Shneiderman, B. . 2002. *Leonardo's laptop: Human needs and the new computing technologies*. Cambridge, MA: MIT Press.

[69] Stoll, Klaus, and Michel Menou. 2003. Basic principles of community public internet access point's sustainability. In *Community networking and community informatics: Prospects, approaches and instruments*, edited by M. Gurstein, M. Menou and S. Stafeev. Saint Petersburg.

[70] Toyama, Kentaro, K. Kiri, D. Menon, J. Pal, S. Sethi, and J. Srinivasan. 2005. PC kiosk trends in rural India. In *Freedom, Sharing and Sustainability in the Global Network Society Conference*. University of Tampere, Finland.

[71] Toyama, Kentaro, Raj Reddy, and Anna Saxenian. 2006. An Introduction to the Best ICTD 2006 Conference Papers. *ITID* (1), http://www.mitpressjournals.org/doi/pdf/10.1162/itid.2007.4.1.1.

[72] Universidad de los Andes. 2009. Public access to information & ICTs final report: Colombia. Seattle: University of Washington Center for Information & Society (CIS).

[73] Unwin, Tim, ed. 2009. *ICT4D: Information and Communication Technology for Development*. Cambridge: Cambridge University Press.

[74] van Dijk, A.G.M.J. 2005. *The deepening divide: Inequality in the information society*. Thousand Oaks, CA: Sage.

[75] ———. 2006. *The network society*. 2nd ed. Thousand Oaks, CA: Sage.

[76] Vanderheiden, G. 2000. Fundamental principles and priority setting for universal usability. *CUU 2000: Proceedings on the 2000 Conference on Universal Usability*:32-38.

[77] Voelcker, Marta. 2008. Public access to information & ICTs final report: Brazil. Seattle: University of Washington Center for Information & Society (CIS).

[78] Walkinshaw, Brady P. 2007. Why Do Riecken Libraries Matter for Rural Development? A Synthesis of Findings from Monitoring and Evaluation. Riecken Foundation, Wash. D.C.

[79] Wanasundera, Leelangi. 2008. Public access to information & ICTs final report: Sri Lanka. Seattle: University of Washington Center for Information & Society (CIS).

[80] Warschauer, M. 2003. *Technology and Social Inclusion: Rethinking the Digital Divide*. Cambridge, MA: MIT Press.

[81] Whyte, Anne. 2000. Assessing Community Telecentres: Guidelines for Researchers. Ottawa: International Development Research Centre. http://www.idrc.ca/en/ev-9415-201-1-DO_TOPIC.html (accessed July 3, 2008).

[82] Wilson, Ernest J. 2004. *The Information Revolution and Developing Countries*. Massachussets: The MIT Press.

[83] Worcester, L., and L. Westbrook. 2004. Ways of Knowing: Community Information-Needs Analysis. *Texas library journal.* 80:102-107.

[84]

[85]

Community Engagement & Infomediaries:
challenges facing libraries, telecentres and cybercafés in developing countries

Elizabeth Gould
Technology & Social Change Group
University of Washington
Roosevelt Commons Building, Suite 400
1-206-685-4116
eagould@uw.edu

Ricardo Gomez
The Information School
University of Washington
Mary Gates Hall, 310E
1-206-685-1372
rgomez@uw.edu

ABSTRACT

Effective infomediaries and community engagement can produce a successful environment to service information needs for underserved populations. This paper analyzes data from libraries, telecentres and cybercafés in 25 developing countries, to assess how infomediaries and community engagement help support the social mission of venues that offer public access to information and communication technologies (ICT). Our results show that while infomediaries and community engagement are critical to facilitate access to information for underserved communities, cybercafés are thriving as public access venues without very strong infomediaries or community engagement, and yet they are perceived as being well staffed and serving community needs. Telecentres and, in particular, libraries, face a particular challenge to fulfill their social mission in the face of the proliferation of cybercafés: they must provide access to ICT, train their staff to be digitally literate and able to support the ICT needs of their communities, and ensure that their community engagement activities include ICT as part of their tools and services.

General Terms
Human factors, Performance

Keywords
telecentres, libraries, cybercafés, public access computing, infomediaries, community engagement

iConference 2010, Feb 3-6, 2010, Urbana-Champaign, Illinois, USA.

1. INTRODUCTION

Three steps are required to serve a population's information needs: understand the population's culture; include someone in the decision-making process who understands the population; and receive direct input from the population from project inception. Gethering user's input enables and involves them in accessing information and solving their information needs in ways that are personally relevant (Bridges.org, 2009).

In this paper, we report findings from a study of information and communication technologies (ICT) used in public venues such as libraries, telecentres and cybercafés in 25 developing countries. Based primarily on qualitative data collected with a common methodological approach across all 25 coutnries, this study offers insights into how libraries, telecentres and cybercafés use infomediaries and community engagement to fulfill the information and communication needs of the communities they serve.

The remainder of the paper is organized as follows: In the next two sections we present a literature review and a description of the methodology used in this research. This is followed by findings and discussion around two topics: infomediaries and community engagement. The paper concludes with a discussion of the implications of the findings. We also identify questions and issues for further research.

2. LITERATURE REVIEW

Most published research on information and library science is pertinent to ICT access in public libraries, and to a lesser extent, in telecentres. Cybercafés have emerged as an important venue in providing public access to ICT for community information needs, although they have received little study concerning the role of infomediaries and community engagement in the success of ICT in public access venues for community development.

In a broad literature review on ICT impacts, Sey & Fellows (2009) pointed out that infomediaries "have been found to be important contributors to the viability and sustainability of a public access venue". The idea is not new. In his study of knowledge and information systems of urban poor, Schilderman (2002) suggested "social networks are the foremost source of information of the urban poor." The poor tend to believe people they trust rather than perhaps more informed contacts with which they do not have close ties. He identified successful ways to meet information needs of urban poor, including involvement of the poor themselves as equal partners, building on local knowledge, use of community-based communication methods, and building the capacity of community based organizations and key individuals within them.

We use the term *infomediary* in a similar way to gatekeepers (Metoyer-Duran, 1993), key informants (Schilderman, 2002), lay information mediaries (Abrahamson & Fisher, 2007) or boundary spanners (Mason, 2003). These authors utilize these terms to refer to a liaison or broker between an individual or group of persons and a group or source of information. We prefer *infomediary* to emphasize the role of brokering or transferring information in a culturally appropriate manner, by taking into account the norms of each group of people they connect.

Community engagement is the ability of community members to work together to achieve shared goals. Bieber et al. (2007) discussed community engagement with three main activities: 1) defining the community; 2) collecting new or existing information in collaboration with community members; and 3) assessing the community's capacity. Similarly, Ballantyne (2002) suggests "foreign content must be matched by the expression and communication of local knowledge that is relevant to local situations… Local content is the expression of the locally owned and adapted knowledge of a community – where the community is defined by its location, culture, language, or area of interest." He emphasized the need for infomediaries to "adapt and synthesize" the information "so that the external content is translated, transformed, and adjusted to suit local situations." An important aspect of community engagement is the ability to produce local content, as discussed by Talyarkhan et al.

3. METHODOLOGY

Our global study needed a common structure and approach to data collection for comparability of results, While retaining flexibility to adapt the research process to the needs and possibilities of each specific context.

Selection of countries

Of 237 possible countries and territories in the world, the final 25 countries (Algeria, Argentina, Bangladesh, Brazil, Colombia, Costa Rica, Dominican Republic, Ecuador, Egypt, Georgia, Honduras, Indonesia, Kazakhstan, Kyrgyzstan, Malaysia, Moldova, Mongolia, Namibia, Nepal, Peru, Philippines, South Africa, Sri Lanka, Turkey, Uganda) went through a selction process that used four successive sets of criteria to focus on a sample of developing countries with a mid-size geography and population, and with existing public library systems. The criteria for country selection were based on size, population and other demographic data[1], degree of freedom of expression[2] and political unrest[3], a measure of "needs and readiness" criteria[4], regional representation, and quality of country research teams. For a more detailed description of the country selection process and rationale see Gomez (2009).

Research Framework

An iterative research design was conducted in two phases. The emergent insights and discussions from Phase 1 guided and sharpened the focus of Phase 2. From the outset, we identified a framework – Real Access – developed in South Africa by Bridges.org. We adapted and refined Real Access, resulting in the Access, Capacity and Environment (ACE) Framework, and structured it as to help understand the range of economic, political, educational, infrastructure, cultural, organizational, and other factors that affect the way people use ICT in public access venues. The three pillars of this framework are: **equitable access**: physical access, suitability, and affordability of the venue, technology access; **human capacity**: human capacity and training (users and staff), meeting local needs, social appropriation; and **enabling environment**: socio-cultural factors, political will and legal and regulatory framework, popular support.

[1] Size (exclude largest and smallest), population (exclude countries with population less than 1 million, and exclude highest population (India, China)), per capita income (exclude countries with per capita income over $11,116), human development index (HDI below 0.5)

[2] Based on Freedom House index: http://www.freedomhouse.org.

[3] Based on U.S. Dept. of State travel advisories.

[4] **Needs criteria**: Income inequality based upon Gini index (2006) from United Nations Development Program; ICT usage: based upon CIA World Factbook (2007); ICT cost: based upon International Telecommunications Union's World Information Society Report (2006). **Readiness criteria**: Politics: based upon World Economic Forum Global Information Technology Report (2006), Transparency International (2007), World Bank Worldwide Governance Indicators (2006); Skills: based upon International Telecommunication Union opportunity skills index (2007); ICT infrastructure: based upon International Telecommunication Union opportunity network index (2007).

Data Collection

Nineteen local research teams were chosen following an international call for proposals. Each team conducted research in local languages, using document reviews, expert interviews, site visits, user surveys, operator interviews, and, in some cases, additional data gathering activities. Detailed country reports were prepared by each research team using a 70-page data-collection template to answer detailed questions about Access, Capacity and Environment issues in each type of venue studied. The use of a common research design and methodology helped make data more comparable, even though the specific ways in which data was collected varied from one country to another in order to make it more locally relevant.

Data Analysis

After careful reading of all reports, we did a detailed annotation of success factors as they were represented in the data. During a facilitated workshop and several group discussions, findings were analyzed, grouped, and categorized, which led to the formulation of the key factors described in this paper.

Limitations of this Study

This study is groundbreaking in its breadth and scope; no other studies have systematically looked at different types of public access venues across multiple countries. Nonetheless, the breadth of the study also means that it does not provide an in-depth analysis of a particular venue, country, or experience, and findings cannot be easily generalized without a clear understanding of the specific context and the analytic framework used.

While the flexibility to translate and adapt the data collection tools to the needs and requirements of each country makes the study more locally appropriate, variations in the way data was collected or presented also makes the comparison of results across countries more problematic. The details discussed here may not be an exact reflection of any single country, but combined across all 25 countries the results represent a meaningful source of trends and patterns about success factors for public access ICT venues.

4. FINDINGS & DISCUSSION

A review of the 25 countries showed public access venues that were most successful at meeting local information needs of underserved communities often contained one or both of two important features: strong infomediaries and/or strong community engagement. These features are experienced differently in each type of venue.

Our findings must be placed in the context of the relative number of each type of venue studied: across all 25 countries, only 12% of the venues are libraries, and 12% are telecentres. Cybercafés account for almost 75% of all public access venues (the balance is a very small proportion of "other" types of venues), according to the results of this study[5]. Accordingly, in terms of public access, the relative weight of cybercafés is three times higher than that of both libraries and telecentres combined. These numbers are important to keep in mind when making programming decisions in both libraries and telecentres, which tend to have stronger infomediaries and community engagement activities than cybercafés.

While infomediary work is generally considered a strong component of the library services, i.e., the role of librarians for helping users find information, libraries tend to have more limited ICT services and their staff is generally not well trained to use ICT tools when they are available. Users put more value in the infomediary role of telecentre and cybercafé operators than that of librarians, because the former are perceived to offer more effective help with ICT tools and services.

Community engagement is a strong feature of community libraries and telecentres. Although community engagement as a purposeful activity is virtually nonexistent among cybercafés, these venues tend to be perceived as meeting local needs more effectively than libraries. A common exception to the popularity of cybercafés is found where there is strong community orientation, ownership and management of the community libraries, as in Argentina, Uganda and Nepal.

Infomediaries in public access venues

Infomediaries can be formal or informal liaisons between communities. A formal infomediary might be a librarian or telecentre/cybercafé operator who has a paid position within the venue. Their job is to reach out to an underserved community, perhaps providing language bridges, literacy connections, needs links, or leadership associations. Equally important are informal infomediaries who may supply similar links, but through different means (e.g., a child to a parent or vice versa, language translators, or unofficial connections between communities). Infomediaries can act on multiple levels: at a community level, between communities (or a community and a venue); as well as at an individual level, between a user and technology. In this paper we focus specifically on formal infomediaries, and we contrast their

[5] While the numbers for libraries tend to be fairly reliable, the numbers for telecentres and, especially, cybercafés are estimates by local researchers. Estimates about number of cybercafés vary widely and are extremely difficult to corroborate, given their informality, lack of organizing body, and quick turnover.

role in libraries with their role in telecentres and cybercafés.

Libraries:

Of the libraries studied, 44% do not offer ICT access to the public. Libraries that do offer ICT access generally do not have digitally literate librarians (trained to use or help users with ICT tools). These factors were prevalent in the majority of the libraries studied, which strongly influences users' negative perceptions about the utility of libraries to meet community needs. This also created negative perceptions concerning the utility of the skills offered by librarians to act as infomediaries for members of the community. We found this 'digital gap' in libraries is successfully bridged when libraries are proactive in meeting community information and communication needs. When libraries are successful at becoming active social and community resource centers, the 'digital gap' of infomediaries in libraries is less apparent than when libraries are only venues that provide access to books and other non-digital resources.

Our study documented numerous practices of infomediaries in libraries in the countries we studied. The following are some typical examples of successful infomediaries in libraries.

Sixty Riecken Foundation Libraries in Central America operate in Honduras and Guatemala. These privately funded libraries began as democratically run community centers for local involvement. Their hallmark is community participation in both the set-up and sustainability of the venues. Because the communities are involved and locals are on the board of directors, emphasis is placed upon local needs and the long-term goals of the community. Library board members develop mission statements, hold elections, and establish library policies. The libraries function as places for people to gather, providing information, entertainment, and socialization. One of the librarians must be an educator from the community. The libraries focus on providing support to people who don't know how to use the information sources, which requires the librarian to act as an infomediary (Arias & Camacho Jiménez, 2008a).

Two publicly-funded information centers in Sri Lanka use intermediaries to disseminate information to groups that lack information literacy: Vidatha Resource Centres and Rural Agricultural Knowledge Centres. These centers disseminate content generated by national research institutes that help to improve quality of life for low-income families, through an increase in productivity and income. "Leadership was a critical factor in the success or failure of venues in achieving their primary objective of meeting the information needs of the communities they serve. Operators that had superior leadership qualities had overcome resource constraints to a great extent and their innovativeness had drawn the communities to the venue to use its service. It was seen that empathy, adopting a participatory approach to the development of the venue, establishing a mechanism to get feedback from the community, and forging links and cooperating with agencies that generate information that is vital for the people are factors that contribute to success in meeting the information needs of the people" (Wanasundera, 2008).

In Uganda, non-profit organizations and foreign agencies established community libraries that target particular sections of the community by providing space for meetings and socialization. The librarian lives in the community and identifies and provides for local information needs in a way best suited to the users. For example, community libraries that serve the primarily rural agricultural communities in Uganda, collect agricultural literature from NAADS (National Agricultural Advisory Services) for local distribution. (Ndaula, 2008).

While the above are fairly typical, a unique example of a national library that serves the community with ICT despite its limited infrastructure is the National Public Library in Tegucigalpa, Honduras. Because computers in this library don't have Internet access, the Library Chief conducts Internet searches on her home computer and presents the information to the users on the following day. She files copies of frequent requests in the library archives in order to respond to similar questions in the future (Arias & Camacho Jiménez, 2008b)

Despite these examples, the users across all 25 countries have a somewhat negative perception of the role of librarians as infomediaries. This trend is exacerbated by growing interest in accessing ICT in public venues. Libraries have comparatively less ICT infrastructure, connectivity and digitally trained staff than cybercafés and telecenres. Despite a long tradition of library information services to the public, with trained librarians and staff (limited as this training may be in most contexts we studied), libraries are perceived as falling behind both telecentres and cybercafés as public access venues offering meaningful infomediary service.

Telecentres:

Telecentre operators are well documented as infomediaries (Benjamin, 2000; Bossio, 2004; Delgadillo, Gomez, & Stoll, 2002; Gomez & Hunt, 1999; Jensen & Esterhuysen, 2001; Parkinson, 2005; Proenza, Bastidas-Buch, & Montero, 2002; Rajalekshmi, 2007), yet we found little evidence of successful telecentres infomediaries, especially when compared to reports about library infomediaries. The few reports about infomediaries in telecentres tend to be positive,

particularly with agriculture or health information. Typical examples telecentres infomediaries follow.

The Telecentros de Porto Alegre program coordinates 35 ICT access points in the region of Porto Alegre, Brazil, are located in community centers and encouraging partnerships with local communities. The facility and utilities are paid for by the community center, while the government pays local youth to provide assistance and maintenance for the computers and Internet. The collaborative effort of government and community contribute to the program's continued success (Voelcker, 2008). Other organizations facilitate a youth training program, also contributing to serving community interests.

Similar projects are successful in the Dominican Republic, where local youth are involved (Alfaro, Molina, & Camacho Jiménez, 2008). The Knowledge and Communication Community Center in Morocelí, Honduras acts as both an ICT access center and a gathering place for youth. The center "offers workshops with every member of the community in mind" (Arias & Camacho Jiménez, 2008b). Because many members of the community don't know how to use the Internet, they often ask children who visit the Center for help, encouraging youth to act as infomediaries. The researchers emphasize the importance of mentoring to enhance information approachability in these venues, helping users understand how the venues are personally applicable (Arias & Camacho Jiménez, 2008b).

The Pallitathya Kendra telecentre in Bangladesh provides infomediaries who travel through the community collecting questions (mobile infomediaries). Back at the center they consult professionals and the content database, then provide answers to inquiries. The infomediary also supplies feedback to the parent telecentre office, which helps to add new content and improve quality (Development Research Network (D.Net), 2008)

Cybercafés:

Cybercafés relceive little attention in research literature compared to libraries and telecentres, particularly in relation to the role of infomediaries. In our study there is little evidence that cybercafé operators are successful infomediaries. Nonetheless, a few exceptional cybercafés are worth mentioning.

In Algeria, as other countries where gender roles are sharply differentiated by religion and social restrictions, female operators are preferred and more trusted as infomediaries. Local researchers reported the majority of the users "said that they are satisfied by the cybercafé especially because Faiza [the telecentres operator] is a very sympathetic lady, there is no stress and there is a print service. They also insist on a fact that Faiza is suggesting a guide of websites. When we ask them to mention all factors that motivate them to use a cybercafé they answer: 'the manager is a lady.' (Bakelli, 2008).

Researchers in Costa Rica report that in cybercafés "it is common to find youths exchanging information about picture uploading, music downloading, templates, layouts, and other tools related to Web 2.0. Although many people do not have the capacities to fully utilize the ICT tools offered in cybercafés, other users and operators help to develop their capacities at least in basic issues (such as e-mail, chat, and information download)" (Sanchez González & Camacho Jiménez, 2008, p. 114).

Cybercafés tend to digitally-literate staff that help users with basic ICT needs. Even if this support is limited, it is valued by users. Across all countries, we found cybercafés are perceived to have higher staff ICT capacity than telecentres and, especially, libraries. Cybercafés are market-driven, therefore more inclined to meet user needs. As specialists in ICT tools and connectivity they tend to support ICT tools and services. Cybercafés are not necessarily driven by a social mission, in contrast to telecentres and libraries, so users' expectations of infomediaries in cybercafés may be far lower than in other public access venues.

In sum, public access venues tend to offer many of the features described by Schilderman for infomediaries: (1) capacity to provide information in an accessible format; (2) willingness to share information; (3) ability to get hold of information and adapt it to a local context; (4) experience, education, knowledge and reliability; (5) accessibility, proximity and helpfulness; (6) social sensitivity and capacity to involve residents; and (7) leadership qualities, influence and moral authority (Schilderman, 2002). Nonetheless, users of ICT in public access venues seek support to use ICT tools and services; this support is offered more effectively in cybercafés. Libraries and telecentres maintain a social development mission, and their staff offers, or is expected to offer, more complex infomediary services, in line with the attributes described by Schilderman. Libraries have a bigger 'digital gap' to fill, and users perceive them as the venues with the least staff capacity, training and disposition to meet local needs.

Community engagement in public access venues

Community engagement of the local population determines the content and services an information venue provides, ensuring the local needs and priorities are addressed.

Our study shows the importance of community involvement for the success of public access venues, especially in community libraries and in telecentres. While cybercafés have few proactive community engagement plans, telecentres and community libraries

often engage community stakeholders in the definition, management and direction of the venues.

According to researchers in Costa Rica, positive results for public access venues depend on increasing human investment by analyzing community information processes, understanding how and why people look for information, the processes involved, who is involved, and what practices are used. (Sanchez González & Camacho Jiménez, 2008)

Argentina illustrates successful community engagement in what are called popular libraries. These libraries are a unique feature in Argentina. They were created by associations of individuals with a dual support system: citizen participation and the governmental Protective Commission of Popular Libraries (CONABIP), which helps create and maintain these institutions. Researchers in Argentina considered both public libraries (centrally supported and funded) and popular libraries in their analysis of libraries as public access venues in that country (Rozengardt & Finquelievich, 2008, p. 71).

Other community based organizations succeed where particular topics or livelihoods bring a community together. The people of the Huaral valley of coastal Peru depend upon water resources to support their agricultural livelihoods. Water resource management and irrigation infrastructure was developed by a small farmers' organization with the help of a local telecentre, which helped to install eleven agrarian information system telecentres in rural communities. The web-based system provides information on water management and cultivation monitoring in the Huaral valley and surroundings. This local community based organization was crucial to achieve success and sustainability. The community helped shape the project, adapt it to the changing environment, and influence policy makers. This project is now being replicated in other valleys in coastal Peru (Bossio & Sotomayor, 2008b).

In South Africa the AIDS/HIV community centers address important community health concerns and build strong community engagement that strengthens their role as public access venues. The centers are near target communities, such as orphans, vulnerable children, or people affected with HIV/AIDS. Many community-based organizations evolve within affected communities. Given the focus on HIV/AIDS, the information tends to be specific and relevant to the needs of the target communities. No access fees are charged at the centers. Information intermediaries are often needed to bridge the gap between technology usage and disadvantaged communities (James, 2008).

Nepal encourages rural communities to become involved and create centers for literacy and social empowerment. Community owned libraries function as social centers and community resource centers. In these venues, community members gather to discuss different issues ranging from civil liberties to human rights. Programs are conducted on health awareness, community development and empowerment.

5. CONCLUSION

Infomediaries and community engagement are critical factors in the success of public access venues offering ICT tools and services with a community development orientation. The community development orientation is an important distinction for libraries and telecentres, which have stronger infomediaries and community engagement. Commercial cybercafés are more numerous but have fewer infomediaries and lack community engagement as part of their mission.

Infomediaries in libraries and telecentres in the countries we studied play an important role in helping to provide and share information in ways that is accessible and useful in the local context, and have a level of education, credibility and helpfulness valued in local communities. Infomediaries in libraries tend to be part of a 'digital gap', as libraries lag behind in offering public access to ICT. Library staff were often unprepared to use or offer support in ICT use in the majority of the public libraries offering ICT. Users increasingly seek ICT access in public access venues, and the 'digital gap' shapes a perception of libraries with the least staff training, preparation and disposition to help meet local needs.

Cybercafés, on the other hand, play a simpler role than libraries or telecentres. In cybercafés operators are expected to serve as infomediaries, and help users with basic ICT use. Cybercafés tend to fulfill this expectation quite well. Thus, users perceive cybercafé staff as skilled and helpful with local needs.

Community engagement, on the other hand, is a critical component of the success of community libraries and of telecentres. Through effective community engagement, these venues become active hubs at the center of community life and information needs. When successful, these public access venues are truly owned and managed by the community they serve, and they become an integral part of local development and transformation. Nonetheless, libraries tend to carry an 'aura of irrelevance' to today's information needs, while cybercafés tend to enjoy an 'aura of relevance' that, combined with their superior numbers, gives users of cybercafés a strong sense of meeting community information needs.

The implications of these findings are threefold:

(1) In addition to offering information services, libraries need to reduce the 'digital gap' in relation to telecentres and cybercafés, by offering

public access to ICT, and transform their 'aura of irrelevance' in public perception.

(2) Libraries need to complement the infomediary skills of librarians and staff with digital literacy skills in order to offer ICT support and assistance. This will come closer to satisfying the information needs of communities that are shifting to ICT as communication tools and sources of information for their community needs.

(3) While libraries and telecentres are not the main source of public access to ICT, they are the main source of relevant infomediaries and community engagement. The social mission of libraries and telecentres would benefit if these venues strengthen their ICT infrastructure and services, but more so if their infomediaries are digitally literate and their community engagement takes full advantage of ICT tools and services.

More research is needed to assess opportunities for libraries and telecentres to collaborate with cybercafés, especially by offering them the support of digitally trained staff to perform important infomediary functions, and by assuring ICT services available in cybercafés are more meaningful to ensure community engagement, and help solve problems.

6. REFERENCES

[1] Abrahamson, J., & Fisher, K. E. (2007). What's past is prologue: Towards a general model of lay information mediary behaviour. *Information Research, 12*(4).

[2] Alfaro, F., Molina, J. P., & Camacho Jiménez, K. (2008). *Public access to information & ICTs: Dominican Republic*. Seattle: University of Washington Center for Information & Society (CIS).

[3] Arias, M., & Camacho Jiménez, K. (2008a). *Public access to information & ICTs: Honduras.* Seattle: University of Washington Center for Information & Society (CIS).

[4] Arias, M., & Camacho Jiménez, K. (2008b). *Public access to information & ICTs: Honduras.* Seattle: presented by Sulá Batsú to University of Washington Center for Information & Society (CIS):.

[5] Bakelli, Y. (2008). *Public access to information & ICTs final report: Algeria*. Seattle: University of Washington Center for Information & Society (CIS).

[6] Ballantyne, P. (2002). *Collecting and propagating local content development*: UK Department for International Development.

[7] Benjamin, P. (2000). Telecentre 2000 Report 1: Literature Review: LINK Centre, P&DM, WITS University.

[8] Bieber, M., McFall, B. S., Rice, R. E., & Gurstein, M. (2007). Towards systems design for supporting enabling communities. *Journal of Community Informatics, 31*(1), 36 pp.

[9] Bossio, J. F. (2004). *Social Sustainability of Telecentres from the Viewpoint of Telecentre Operators: A Case Study from Sao Paulo, Brazil.* London School of Economics, London.

[10] Bossio, J. F., & Sotomayor, K. (2008b). *Public access to information & ICTs final report: Peru.* Seattle: presented by Alfa-Redi to University of Washington Center for Information & Society (CIS):.

[11] Bridges.org (2009). 12 Habits of Highly Effective ICT-Enabled Development Initiatives Retrieved from http://www.bridges.org/12_habits

[12] Delgadillo, K., Gomez, R., & Stoll, K. (2002). Community telecentres for development : lessons from community telecentres in Latin America and the Caribbean: IDRC, Ottawa.

[13] Development Research Network (D.Net) (2008). *Public access to information & ICTs final report: Bangladesh.* Seattle: presented to University of Washington Center for Information & Society (CIS):.

[14] Gomez, R. (2009). Structure and Flexibility in Global Research Design: Methodological Choices in Landscape Study of Public Access in 25 Countries. University of Washington.

[15] Gomez, R., & Hunt, P. (Eds.). (1999). *Telecentre Evaluation: A Global Perspective*. Ottawa: IDRC.

[16] James, T., et al. (2008). *Public access to information & ICTs final report: South Africa*. Seattle: University of Washington Center for Information & Society (CIS):.

[17] Jensen, M., & Esterhuysen, A. (2001). *The Telecentre Cookbook for Africa: Recipes for self-sustainability*. Paris: UNESCO.

[18] Mason, R. M. (2003). Culture-Free or Culture-Bound? A Boundary Spanning Perspective on Learning in Knowledge Management Systems. *Journal of Global Information Management, 11*(4), 20-36.

[19] Metoyer-Duran, C. (1993). *Gatekeepers in Ethnolinguistic Communities*. Norwood, NJ: Ablex Publishing.

[20] Ndaula, S. (2008). *Public access to information & ICTs final report: Uganda*. Seattle: University of Washington Center for Information & Society (CIS):.

[21] Parkinson, S. (2005). *Telecentres, Access and Development: Experience and Lessons from Uganda and South Africa*: IDRC.

[22] Proenza, F., Bastidas-Buch, R., & Montero, G. (2002). Telecenters for Socioeconomic and Rural Development in Latin America and the Caribbean. Inter-American Development Bank. 17. Retrieved from http://www.iadb.org/sds/itdev/telecenters/exsum.pdf

[23] Rajalekshmi, K. G. (2007). E-governance services through telecenters: The role of human intermediary and issues of trust. *Information Technologies and International Development, 4*(1), 19-35.

[24] Rozengardt, A., & Finquelievich, S. (2008). *Public access to information & ICTs final report: Argentina*. Seattle: University of Washington Center for Information & Society (CIS):.

[25] Sanchez González, A., & Camacho Jiménez, K. (2008). *Public access to information & ICTs final report:*

Costa Rica. Seattle: University of Washington Center for Information & Society (CIS).

[26] Schilderman, T. (2002). *Strengthening the knowledge and information systems of the urban poor*: Dept. for International Development (DFID).

[27] Sey, A., & Fellows, M. (2009). *Literature Review on the Impact of Public Access to Information and Communication Technologies*. Seattle: Center for Infromation & Society, Univ. of Washington.

[28] Voelcker, M. (2008). *Public access to information & ICTs final report: Brazil*. Seattle: University of Washington Center for Information & Society (CIS).

[29] Wanasundera, L. (2008). *Public access to information & ICTs final report: Sri Lanka*. Seattle: University of Washington Center for Information & Society (CIS).

[30]

Conferences, Community, and Technology: Avoiding a Crisis

Jonathan Grudin
Microsoft Research
One Microsoft Way
Redmond, WA USA
+1 425 706 0784
jgrudin@microsoft.com

ABSTRACT
Computer Science in North America has embarked on a course unique in academic scholarship. It has turned conferences into repositories of polished work, little of which ever evolves into journal articles. Senior researchers feel that the conferences are in crisis. I consider the origins and consequences of the shift to conferences, concluding that it has led to an evolutionary cul-de-sac that the Information field would do well to avoid. The crisis is described as centered on reviewing, but it is at heart a crisis of community.

Categories and Subject Descriptors
K.2. History of Computing
K.7.2. The Computing Profession: Organizations.

General Terms
Management, Design, Human Factors.

Keywords
Community, Disciplines, Conferences.

1. THE ECOLOGY OF SCIENTIFIC AND TECHNICAL COMMUNICATION
It is often observed that new technologies are inserted into existing processes to replace older technologies. "Design a word processor with the typewriter as a model," we were advised. Then, over time, the processes and the technologies are restructured. The goal is to improve upon the status quo, but long-term consequences can be unpredictable.

My topic is the effects of new technologies on the processes of research dissemination. We inserted transformational technologies into a complex ecology of books, journals, and conferences with minimal reflection on what would eventually replace the "iron horse" stage. Email replaces informal conversations and phone calls, word processing is used to prepare articles, PowerPoint replaces slides, authors put articles on the Web or blog their findings. We know this is just the beginning. What are the destinations, and how will we reach them?

Between 1997 and 2003, when I was Editor in Chief of *ACM Transactions on Computer-Human Interaction*, the Internet and the Web were perceived to threaten the existing business models of publishers. My 2004 essay *Crossing the Divide* outlined goals of scholarly communication shown in Table 1 [5]. The goals frequently conflict. Careful reviewing trades off against rapid dissemination. The goal of archiving all useful results can run up against page and cost constraints. A submission that an author considers to be an original contribution, an editor may declare to be out of scope. This creates a complex force field that drives books, journals, conferences, and workshops to different niches, each niche representing a different weighting of priorities. The resulting landscape varies across disciplines, and within a discipline can vary by country or continent, marked by differences in the nature of the scholarly activities, the approaches to assessing contributions, legacy practices and traditions, and the state of digital technology integration.

Digital technologies have affected scholarship in diverse fields. Physics and Mathematics are frequently-cited examples. My focus is on lessons for Information from the Computer Science experience

2. A CRISIS ENGULFING COMPUTER SCIENCE
In 2009, four essays in three issues of *Communications of the ACM* [1, 2, 9, 10] addressed "a growing crisis" in the computer science community. They argue that a focus on conference publication has led to deadline-driven, short-term research at the expense of journal publication, a reviewing burden that drives off prominent researchers, and high rejection rates that favor cautious incremental results over innovative work. In one essay, Ken Birman and Fred B. Schneider observe that in Computer Science, "in the past, journal publications were mandatory for promotions at leading departments. Today, promotions can be justified with publications in top conferences." [1] The resulting deluge of conference submissions creates reviewing challenges.

Although Birman and Schneider focused on reviewing, other of the interlocking Table 1 goals arose. Conference deadlines insure timely dissemination of results, but undermine originality, as novel papers are "time consuming to read and understand, so they are the most likely to be either completely misunderstood or underappreciated." More submissions lead to fewer broad program committee discussions and lower quality reviews. Birman and Schneider describe a "death spiral" in which senior researchers cease participating in review panels.

A *CACM* reader eagerly anticipating solutions may be disappointed. Birman and Schneider recommend (i) returning journals to prominence, a plea echoed by other commentators, and (ii) giving authors of conference submission no feedback, either to discourage premature submission or to reduce reviewer workload. In short, a plaintive call for an unlikely return to the past, lacking analysis of why Computer Science in the United States shifted to conference publication in the first place. It did not happen to Computer Science in Europe or Asia, or in other competitive, quickly-evolving fields such as Neuroscience or Physics.

Function	Time	Goal
Production	Venue creation	Defining scope.
		Defining quality or soundness.
		Defining originality.
	Reviewing	Measuring the value of submissions.
		Helping authors improve submissions.
	Publishing	Disseminating results quickly.
		Distributing results broadly.
		Archiving and providing access to all useful results.
		Publishing on schedule / maintaining content flow.
		Adhering to page count constraints.
		Making or not losing money.
Group well-being	Long-term effects	Growing and maintaining a research community.
Member support		Helping individual community members succeed.

Table 1. Goals of journals, conferences, and workshops. (Based on [5].)

3. A DISRUPTIVE TECHNOLOGY AND PROCESS: WORD PROCESSING AND ARCHIVED PROCEEDINGS

Major changes generally have multiple causal factors. In this case a new technology and a related decision seemed to transform the situation like a key opening a lock. The use of text editors and word processors by computer scientists enabled timely, inexpensive production of presentable conference proceedings. This perturbed the complex ecology, setting in motion a series of adjustments that are still being worked out.

Prior to the 1980s, the rare proceedings available at a conference required expensive editing and typesetting or typewritten pages with figures and tables pasted in. By the early 1980s, most computer science researchers had access to text editors or word processors, graphics packages, and printers that supported standard font sets. Conferences published formatting instructions for final versions that were intended to yield a consistent look. Text processors of the 1980s had limited formatting capability, so proceedings lacked today's uniformity, but they looked decent enough. Costs were contained by having authors do most of the work and by shifting from hardcover to trade paperback format.

CHI conferences had inexpensive proceedings prepared in advance from the beginning in 1983. The first international HCI conference, INTERACT 1984 in London, first produced a two-volume provisional paperback proceedings available on site, then a single-volume hardcopy proceedings with a more uniform look. From 1985 on, few if any major CS conferences produced proceedings after the event.

This technology change was not disruptive by itself. The second factor was the existence in North America of a non-profit professional organization that served computer scientists and organized many major conferences. The Association for Computing Machinery (ACM) saw an opportunity in proceedings of trade paperback quality appearance, low production cost, and very low per-copy incremental printing cost. ACM printed many more copies than there were conference attendees and set about marketing the surplus to libraries, the lifeblood of technical and scientific publishers.

In addition, some conference-cosponsors, such as the Special Interest Group on Computer-Human Interaction (SIGCHI), sent copies as a benefit to their thousands of members. Finally, and perhaps of greatest significance, mail-order copies could be purchased very inexpensively by anyone, years later. With library uptake slow, there was effectively an inexhaustible supply.

As a result, fifteen years before the digital library, ACM conference proceedings were archived and widely accessible. These were the two original purposes of journals! The ecological balance of technical communication in was disrupted, with effects that are still being worked out a quarter century later

4. UNINTENDED CONSEQUENCES (1): JOURNAL DECLINE

As editor of *ACM TOCHI*, I was frequently told by senior colleagues that they considered journals irrelevant. This is a development that the commentators lament. Why did it happen?

The short answer is that the major players had incentives to sharply drive up the quality of conference papers. This reduced the incentive for continuing to improve the work and raised the bar substantially for those who tried to do so.

To sell proceedings to libraries, ACM had a stake in papers being of the highest possible quality. In addition, libraries were more likely to acquire and shelve thinner volumes. Many authors, when they realized that conference papers would be immortal, desired to make a good impression. Also, when Digital Libraries, site licenses, and Internet access arrived, academic Computer Scientists realized that their conference product could be easily viewed and judged by colleagues evaluating job candidates, tenure cases, or promotions. A self-policing function arose: If we

let the quality waver, we could lose hard-won respect from peers in other disciplines.

Consider CHI as an example. A 1982 conference led to its formation. Proceedings for that conference were not available after the event. It had a 45% acceptance rate. Over CHI's first 13 years, the median acceptance rate was 27%, the maximum 39%. For the next thirteen years, the median was 22% and never exceeded 25%. For three successive years it was 15%-16%. The 25% ceiling that has held since 1995 coincides with the rise of HCI within academic Computer Science. At many U.S. universities, computer scientists convinced colleagues from other fields to weight papers from very selective conferences highly in appointments and promotions; 25% was a good demarcation.

Assume that the authors of a CHI paper would like to improve it, by responding to reviewer suggestions for which there had not been time or space in the final version, by extending the literature review to actually discuss some of the papers cited, by expanding their own discussion, or by including additional analyses or details. In the 1980s, ACM and IEEE policy was that authors of conference papers, which were not archival, could republish them as journal articles, which were. Usually journals expected more, but many excellent conference papers were republished verbatim or close to it. However, as conference papers developed an afterlife by being effectively archived (and later unequivocally archived in Digital Libraries), IEEE and ACM shifted policy to discourage republication, now considered self-plagiarism.

How much must be changed to merit journal publication? That bar has been steadily raised by publishers, editors, and reviewers. Some even consider the merger of two related conference papers into a single journal article to be unacceptable: A new publication requires previously unpublished data. The result, which I have not seen discussed despite its centrality to the decline of journal publication, is that authors of selective conference papers often find it prohibitively difficult to publish in journals.

Correlational data exists that bears on the above hypotheses, but given space constraints, I will conclude this section with a supportive logical argument. In Europe and Asia, professional organizations did not provide low-cost post-conference access to proceedings. Authors who wished their work to be accessible had to progress it to journal publication. Journals remained the major or only academic measure of achievement. Articles in the leading U.S. HCI journals shifted from being mostly authored by Americans to being mostly authored abroad. Interest in journal impact factors has been high among European computer scientists, but not among Americans -- impact factors were generally not calculated for conferences and did not measure citations appearing in papers in selective conferences.

European and Asian conference acceptance rates generally stayed higher, although some rose under competitive pressure from U.S. conferences. Only recently have I seen growing acceptance of selective archival Computer Science conference papers in some European countries.

In 2004, a prominent UK researcher wrote about CHI:

> HCI's love of conferences is a fluke of history. We all know this. CS in general has suffered from it, but is steadily moving away. CHI however digs in, with more and more death rattles such as CHI Letters. Being conference centred is bad for any field: bad for its archival material, bad for its conferences, and worst of all, really bad for the respect that we command with other communities. SIGCHI needs to move away from bolstering up conference publications. It needs to use journals for journal stuff and conferences for conference stuff. [3]

He was wrong about the direction of Computer Science, and at least premature in diagnosing CHI's expiration. The point, though, is that he saw the problem as an American problem, affecting CHI but not European HCI.

Birman and Schneider decry the erosion of journals and describe a "death spiral" in which people overburdened by deadline-driven conference reviewing cease reviewing for journals. Perhaps in our dynamic field the shelf-life of some results is short or a conference paper captures the essence of the research, but I agree that the additional reflection afforded by an iterative review and revision process is valuable. However, considering the forces that led to the present state, a return to journal preeminence seems unlikely.

Information Schools, comprising computer scientists and researchers from other disciplines, wrestle with the assessment of publication venues—but so do many other schools with CS departments. The iCaucus and Information Conference will have to decide whether a new journal should be formed and whether the proceedings should be archived. However, the crisis that the Information field should work to avoid is not this fifteen-year-old dilemma. It is an emerging second-order effect of the shift to conference publication. The U.S. Computer Science crisis is a crisis of community.

5. UNINTENDED CONSEQUENCES (2): COMMUNITY DECLINE

The core problem confronting Computer Science is that their major conferences focus on assessing and showcasing the field's quality work, a role formerly filled by journals, and have largely abandoned the community-building and community-maintenance function that conferences traditionally fill. In the absence of an effective replacement, there has been a gradual but cumulatively significant decline in the sense of community in major Computer Science sub-disciplines, with no bottom in sight. Diverse factors may be at work, but let's step back to consider a framework from social psychology (Table 2).

	Production	**Group Well-Being**	**Member Support**
Inception	Production demand and opportunity	Interaction demand and opportunity	Inclusion demand and opportunity
Problem-Solving	Technical problem-solving	Role network definition	Position and status attainments
Conflict Resolution	Policy resolution	Power and payoff distribution	Contribution and payoff distribution
Execution	Performance	Interaction	Participation

Table 2. McGrath's Group Functions and Modes [8].

Joseph McGrath identified functions and modes of activities in teams or groups. At different times, groups take on new tasks (inception), work on them (execution), solve problems that arise, and resolve conflicts [8]. Of significance to us are the columns. Groups continuously engage in activities that address production (their *raison d'être*), team health, and member support. We may address the second and third without conscious consideration, but we ignore them at our peril.

Studies of group support technologies tend to focus on the lower left cell, *performance*—effects on productivity, return on investment. Technologies that have positive effects on performance in experiments may founder in practice due to negative effects in other cells. Conversely, technologies that show no short-term performance benefits in studies may have positive effects in other cells that could benefit performance over longer periods [4, 6, 7].

Table 1 is further evidence of a bias toward the production function. It seemed reasonable, yet it focuses overwhelmingly on production. Contrast it with Table 3.

This assignment of function to venue omits considerable nuance. For example, doctoral symposia or full-day workshops held in conjunction with a conference provide member support. But the broad picture is clear, as is the contrast with other fields. A friend described the annual Neuroscience Conference as a must-attend event: "It is where you find out what is happening!" It has 15,000 presentations and 30,000 participants. Quality is not the point, community is. Journals are where he finds quality; workshops are a source of information and feedback for work in progress.

Highly selective conferences work against many of the group well-being goals in Table 3. When 75% of paper submissions are rejected, it is difficult for researchers from allied fields or new researchers who do not know the conventions to break in. Setting aside the fact that being rejected is generally an off-putting experience, many people must present to get travel funds, so engagement is curtailed. The rejected material becomes fodder for spin-off or sub-group activities, which proliferate, scattering people, their energy investments, and the relevant literature. Community identity declines. For a typical topic, only one in four submitted papers is presented and much work in progress is not even submitted, so the conference is not a place to find out what is happening in one's specialty area. This further opens the door for new or competitive venues.

Membership data for Computer Science special interest groups since 1990 can be found at www.acm.org/sigs. SIGCHI membership peaked in 1992. It fluctuates but is currently down about 20% from that level. Conference attendance peaked in 2001. This is true despite an unquestioned increase in faculty, students, and practitioners focused on HCI. Graduate students are a steadily rising fraction of conference attendees and presenters. Practitioners disappeared from the program and then from the audience. Some member support functions are served—students get visibility and speaking experience, professors get their names on papers whether or not they attend. But for most people, rewards for attending have diminished. The papers can be read in the proceedings. In the early years, papers were assigned discussants, but the polished papers that make it through today's competitive review processes leave less room for comment. This is especially true given the Birman and Schneider observation that original or controversial papers are unlikely to survive the review process. One frequently hears statements such as, "I submitted two papers. The original and interesting one didn't make it. The more boring, incremental paper did."

In ACM SIGCHI, once-active community forums are gone. The newsletter, the *SIGCHI Bulletin*, was vibrant through the 1980s. A market research study in the early 1990s found it was avidly read. The past decade it withered and died. The CHI email distribution list used solely for event announcements was once a lively

Function	CS Venue (then / now)	Goal
Production	**Journal / Conference**	Collecting and distributing research results.
Group well-being	**Conference / Not Clear**	Establishing community identity.
		Developing members, maintaining engagement
		Recruiting new members.
		Interacting with parent and sibling organizations.
		Interacting with competitive or rival organizations.
		Managing subgroups and spin-offs.
Member support	**Workshop/ Workshop**	Helping students get visibility and jobs.
		Helping assistant professors get tenure.
		Helping associate professors get promoted.
		Helping full professors get honors.
		Helping practitioners prosper.
		Recognizing research and service contributions.

Table 3. Goals in U.S. Computer Science and venues before / after the shift.

discussion forum. The web-based CHIPlace forum was a focus of community discussion a decade ago; use trailed off and it was taken down. Business meetings held at the conference were once heavily attended and a source of passionate argument—a petition circulated at one conference forced an election of officers. Today gatherings are poorly attended; complaints over reviewing and heavy-handed program committee members are a major focus. A sense of community is found at the program committee meeting, restricted to a small number of mostly senior people. If pressures to save time and travel expense lead to distributed program committee meetings, social interaction will decline further.

The bottom line is that a conference that rejects 75% of submissions may not fill a community-building role unless it has some other irresistible draw—ICIS and SIGGRAPH can be must-attend venues despite high paper rejection rates due to their links to job interviews and exhibitions, respectively. Otherwise, this path seems problematic.

6. ALTERNATIVE PATHS

Academic Computer Science benefits from the status quo, which is intricately woven into its accreditation process. Change will not be easy. It may not be necessary, although as noted in the CACM commentaries, reviewers are more difficult to enlist. Conferences once run entirely by volunteers now contract out much of the work, but reviewing cannot be outsourced. People whose papers are rejected shift their efforts to the conferences that subsequently accept the papers. Birman and Schneider's solution, 'stop writing useful reviews," does not seem viable given the claims for quality in which we take such pride.

Computer Science in Europe has recently shifted toward greater recognition of selective conference papers. Other fields have not. The role of our strong, non-profit professional organizations was significant. How other fields react as online preservation becomes ever easier is something to watch.

Information Schools have some time to explore options. It seems appealing to stress the community-building focus of most major conferences. I attended an American Anthropological Association meeting—7000 anthropologists! Presentation quality varied, the high energy level did not. But it will be a challenge, particularly given that Information School faculty from the Computer Science community may be unaware that other ways of life are possible.

Those who frequent highly selective conferences expect polished work. They often decide which session to attend on short notice. Attending a larger, less selective conference, many complain bitterly about presentation quality. They do not realize that with an hour or so preparation based on the program and other materials at hand, one can have as positive an experience at a large inclusive conference as at a selective conference. But reeducating the Computer Scientists among us is not the only challenge.

The accreditation process must be considered. A large, inclusive conference could accept 80% of submissions for presentation in parallel tracks and identify 20% as Best Paper Nominations. This could provide a quality measure for those who need one, enable more people to present and learn what is going on in their areas, and help people plan their attendance around strong papers.

Another concern is so-called self-plagiarism. It is not expensive to host 80% of submissions online, but is it all archival? One can attach labels, but people put everything on their CVs. This difficulty is inherent in our increasingly visible world. With no cost to putting drafts online, the issue of multiple versions being published in some form is unavoidable.

Perhaps technology, having helped create the problem, can help fix it. Wikipedia's articles with complete version history and discussion pages are a possible model. The mediawiki software has weaknesses, not least of which is that references are not an object type. But perhaps a researcher at some point registers a draft with a system, controlling access, after which all versions, comments, and reviews are recorded. The work may initially be private, then opened to friends or colleagues, later submitted to a workshop, then conference, and maybe a journal or journal-level accrediting process. At each stage the version history is there, review comments are accumulated, the work develops. Self-plagiarism isn't an issue; anyone can inspect the history.

There are issues, technical and otherwise. What happens when an author combines two or three works into one larger work, or when co-authorship changes? How is copyright managed if I submit to an ASIST workshop, then want to submit a version to a CHI conference, and later to a for-profit publisher's journal?

7. CONCLUSION

Information Schools wrestle with issues of identity, direction, and quality measurement. Computer Science offers one model, which appeals to some because they come from the field or because Computer Science has been successful. But in the view of many, Computer Science is in trouble. Therefore, it makes sense to look closely at the current state and understand how it came to be. It was reached not through planning and consideration of alternatives, but because the field was pushed unwittingly down a path by using technology in an obvious and beneficial way, which nevertheless had unintended consequences.

There is more to say and more questions to ask. Let's continue the discussion.

8. REFERENCES

[1] Birman, K. & Schneider, F.B. 2009. Program committee overload in Systems. *Communications ACM, 52,* 5, 34-37.

[2] Crowcroft, J., Keshav, S. and McKeown, N. 2009. Scaling the academic publication process to Internet scale. *Commun. ACM, 52*, 1, 27-30.

[3] Cockton, G. 2004. Email communication, 22 January.

[4] Dennis, A.R. & Reinicke, B.A. 2004. Beta versus VHS and the acceptance of electronic brainstorming technology. *MIS Quarterly, 28,* 1, 1-20.

[5] Grudin, J. 2004a. Crossing the divide. *ACM Transactions on CHI, 11*, 1, 1-25.

[6] Grudin, J. 2004b. Return on investment and organizational adoption. *Proc. CSCW 2004*, 274-277.

[7] Grudin, J. 2008. McGrath and the Behaviors of Groups (BOGs). In T. Erickson and D.W. McDonald (Eds.), *HCI Remixed: Reflections on Works That Have Influenced the HCI Community.* MIT Press.

[8] McGrath, J. 1991. Time, interaction, and performance (TIP): A theory of groups. *Small group research, 22*, 2, 147-174.

[9] Vardi, M. Y. 2009. Conferences vs. journals in computing research. *Commun. ACM, 52*, 5, 5.

[10] Wing, J.M. and Guzdial, M. 2009. CS woes: deadline-driven research, academic inequality. *Commun. ACM, 52*, 12, 8.

Wikipedia Community Spaces: Comparative Analysis of Behaviors Across Talk Pages in Four Languages

Noriko Hara
School of Library & Information Science, Indiana University
1320 E. 10th street
Bloomington, IN 47405 USA
+1 (812) 855-1490
nhara@indiana.edu

Pnina Shachaf
School of Library & Information Science, Indiana University
1320 E. 10th street
Bloomington, IN 47405 USA
+1 (812) 856-1587
shachaf@indiana.edu

Khe Foon Hew
National Institute of Education, Nanyang Technological University
NIE2-03-25, 1 Nanyang Walk,
Singapore 637616
+011 (65) 6790-3282
khefoon.hew@nie.edu.sg

ABSTRACT
This paper reports a cross-cultural analysis of Wikipedia communities of practice (CoPs). First, it argues that Wikipedia communities can be analyzed and understood as CoPs. Second, the similarities and differences in norms of behaviors across four different languages (English, Hebrew, Japanese, and Malay) and on three types of discussion spaces (Talk, User Talk, and Wikipedia Talk) are identified. These are explained by Hofstede's dimensions of cultural diversity, the size of the community, and the role of each discussion area. This paper increases our understanding of the various Wikipedia communities and expands the research on online CoPs, which have rarely examined cultural variations across multiple CoPs.

Categories and Subject Descriptors
H.5.3 [**Group and Organization Interfaces**]: Collaborative computing

General Terms
Human Factors, Languages.

Keywords
Online communities of practice, norms of behavior, Wikipedia, cross-cultural

1. INTRODUCTION
The proliferation of the social Web and its participatory nature gives rise to many online communities, some of which undertake a common practice that links users together. In this environment, users often become "prosumers," who are consumers and producers at the same time (Tapscott & Williams, 2008). These prosumers are sharing knowledge through massive collaborative efforts, for example, by writing encyclopedic articles on sites such as the Wikipedia, or by providing answers to questions posted on Q&A sites such as Yahoo! Answers. Various companies identified the potential utility of prosumers by soliciting product and research ideas from the prosumers, who are not employed by their organizations (Tapscott & Williams, 2008). Wikipedia users are engaged in knowledge sharing processes by implementing common practices that create and maintain shared identity. This type of knowledge sharing is pertinent to the idea of "communities of practice" (CoPs) (Lave & Wenger, 1991).

While research on online CoPs expand and they are widely implemented by organizations across the globe, only a few cross-cultural analyses of CoPs have been conducted in the past. For instance, Pan and Leidner (2003) studied a knowledge management system that resides within an international organization to connect multiple CoPs, but they did not particularly focus on a cross-cultural analysis of CoPs. There is a need for analytical, conceptual, and comparative work regarding online CoPs. In particular, as organizations become more global and deploy more cross-border CoPs than ever before, there is a need for cross-cultural analyses of CoPs.

The global nature of Wikipedia makes it an interesting case for such cross-cultural analyses; over 75% of Wikipedia is written in languages other than English. Still, despite the multi-lingual nature of Wikipedia (75,000 active members in more than 260 languages (Wikipedia: About)), prior studies are predisposed to investigate only the English version of Wikipedia, with a few exceptions (e.g., Pfeil, Zaphiris, & Ang, 2006). A cross-cultural analysis of Wikipedia as a CoP would inform our understanding of the impact of information and communication technology on cultural differences.

2. BACKGROUND
First, there is a need to examine if Wikipedia communities possess characteristics of CoPs. Then, a brief outline of cultural differences is presented. Only after that, we can contribute to the discussion of cross-cultural analysis of CoPs by analyzing multiple language versions of Wikipedia.

2.1 Wikipedia as Communities of Practice
In this paper we use the following definition of CoPs: "groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis" (Wenger, Dermott, & Snyder, 2002, p. 4). Wenger (1998), in his seminal book, identified four characteristics of communities of practice: practice, community, meaningful learning, and identity. We argue here that all four characteristics of CoPs are prevalent within the Wikipedia community; the manifestation of each of these characteristics in Wikipedia is described next.

1. Practice: Wikipedia users share practice and are engaged in knowledge sharing as seen in discussion spaces (e.g., Stvilia, Twidale, Smith, Gasser, 2008). Users write original articles, edit and improve existing articles, provide quality assurance,

fight against acts of vandalism, participate in policy setting, and are engaged in community building and maintenance activities. Sharing these practices fosters a sense of community among Wikipedia users.

2. Community: Wikipedia users developed a set of community norms and policies. Users are expected to adhere to these norms and the Wikipedia policies (e.g., Riehle, 2006). New members of the community are informed and encouraged to learn these norms of behaviors by seasoned users who have mastered processes, policies, and practices on the Wikipedia. Bryant, Forte, and Bruckman (2005) report that new users tend to focus on editing individual articles and later become motivated to contribute to the well-being of the Wikipedia community and the Wikipedia project as a whole. This indicates that seasoned Wikipedia users develop a sense of community.

3. Meaningful learning (learning in context): Wikipedia users learn to become Wikipedians (e.g., Bryant, et al., 2005); they learn how to behave, how to write, and how to be a member (in good standing) of the community. When examining how new Wikipedia users become Wikipedians, Bryant, et al., (2005) applied a concept of legitimate peripheral participation used in the communities of practice literature (Lave & Wenger, 1991). "Legitimate peripheral participation" is a term used to describe a mode of participation in which newcomers to a community peripherally participate in the practice. Though peripheral, the participation is legitimate in the sense that these new apprentices can observe other members, especially the more experienced, and learn how to become full members of the community.

4. Identity: Shared identity is partially formed around the share practice of the CoP. The Wikipedia user develops an identity as a Wikipedian (Bryant, et al., 2005). Anthony, Smith, & Williamson (forthcoming) contend that, as shown among open source developers, reputations and group identity are one of the motivators to contribute to public goods. Due to the strong group identity on the Wikipedia community, contributions are flourishing.

Because Wikipedia possesses characteristics of CoPs, we are able to contribute to the discussion of cross-cultural analysis of CoPs by analyzing multiple language versions of Wikipedias.

2.2 Culture: Hofstede's Dimensions

Culture is defined as "the collective programming of the mind which distinguishes the members of one group or category of people from another" (Hofstede, 1991, p. 5). One way that scholars try to understand the nature of cultural differences is to understand a pattern composed of a combination of dimensions (Straub, Loch, Evaristo, Karahanna & Srite, 2002). While various such approaches to culture exist (e.g., Hall; 1976; Hofstede, 1991; Trompenaars & Hampden-Turner, 1998), Hofstede's (1991) five dimensions are the most commonly utilized pattern in cross-cultural research. The dimensions include: low/high power distance, individualism/collectivism, masculinity/femininity, uncertainty avoidance, and long-term/short-term orientation. The fifth dimension was a later addition to his framework based on the contribution of Chinese scholars (Hofstede, 2008). Empirical studies that utilize the dimensional approaches, however, report inconsistencies in the context of groups that use technology. Most of the studies focus only on 3-4 dimensions, mostly on Hofstede's (1991) dimensions (Mayers & Tan, 2002), and typically, they only focus on the individualism/collectivism dimension.

2.3 Research Gap and Questions

Prior research on online CoPs is largely constrained within organizational boundaries, defies the examination of explicit norms of behaviors, and lacks extensive cross-cultural analysis. Research on Wikipedia primarily focuses on the English Wikipedia and tends to overlook the Wikipedias in languages other than English. This paper addresses this gap and examines the variation of norms of behaviors among Wikipedia users in four languages and on three Talk pages (Talk, Wikipedia Talk, and User Talk pages). The study aims to answer the following research questions: How do Wikipedia norms of behavior vary across languages? What are the (cultural) variations among Talk pages?

3. METHODS

This section discusses the method of data collection and data analysis procedures. Content analysis of 120 Wikipedia Talk pages in four languages was conducted. Thirty pages were randomly selected from Wikipedia in each of the four different languages—English, Hebrew, Japanese, and Malay; two languages are spoken in Western cultures (English and Hebrew) and two in Eastern cultures (Japanese and Malay). These Wikipedias also vary in their sizes: two large Wikipedias, with over 100,000 articles each (English, with 2,580,417 articles, and Japanese, with 526,800 articles) and two smaller Wikipedias, with less than 100,000 articles (Hebrew, with 83,034 articles, and Malay, with 30,890).

3.1 Data Collection

First, we randomly selected talk pages from the English Wikipedia, for a pilot study. Using the Wikipedia search capabilities, a list frame for the various talk pages was generated on July 2007 to use for the random sampling. Random sampling of Wikipedia pages is not a common practice; prior research on Wikipedia norms used purposeful sampling (e.g., Viégas et al., 2007). The sample in the pilot study included 30 pages, 10 pages from each of the three name spaces, Talk, User Talk, and Wikipedia Talk. These 30 pages include more than 700 posts.

Second, the same sampling procedure used in the pilot study was utilized to collect pages from the other three Wikipedias in September 2008. The sample at this step included 90 additional pages, from the three other languages (Hebrew, Japanese, and Malay). The final sample used in this study included 120 talk pages, 30 talk pages from each of the four Wikipedias and from three different name spaces, 10 from each type of talk page. In total, this includes over 2,700 messages.

3.2 Data Analysis

The coding scheme was developed from the ground up during the pilot study, using interpretive content analysis; the first and second authors coded a subset of the data and suggested a list of codes. These codes were discussed and grouped into broader categories. The discussions with the third author refined individual codes. The final coding scheme includes fifty-six codes under 3 categories: writing norms, information sharing,

and community well-being[1]. This coding scheme was later used to code the pages and to compare norms of behavior among the three types of talk pages.

The data were sorted under each code and frequency tables were created for each language. Each of the authors and an independent researcher coded the selected pages in one language, which was his/her native language. All four coders coded 10% of the data of the English Wikipedia to examine the level of inter-coder reliability (number of agreements, divided by the sum of the number of agreements plus number of disagreements). Inter-coder reliability was high: 93% between the Hebrew and English Wikipedias, 85% between the Japanese and English Wikipedias, and 86% between the Malay and English Wikipedias.

Because the study was designed to investigate the difference between Eastern and Western cultures and larger and smaller Wikipedias, the data were further aggregated based on culture and size. For each of the four conditions (i.e., Western, Eastern, large, and small Wikipedias) we created an aggregated frequency column for Talk, User Talk and Wikipedia Talk. Subsequently, cross tabulation analysis was conducted on the data using SPSS 17.0.

Table 1. Total number of posts in the three namespaces among four languages

	Talk (627)	**User Talk (833)**	**Wikipedia Talk (1253)**
English (1264)	288	421	555
Hebrew (686)	156	333	197
Japanese (567)	162	25	380
Malay (196)	21	54	121

Table 1 presents total number of posts for the three namespaces in four Wikipedias. The English Wikipedia has the largest number of messages in Talk, User Talk, and Wikipedia Talk pages because it is the largest Wikipedia. The size also explains the number of messages posted on Wikipedia Talk pages, because the larger the Wikipedia becomes, the more it involves administrative tasks. However, if we only consider the size of Wikipedias, the Japanese Wikipedia should have accumulated the second most messages, then Hebrew, and Malay in all three namespaces. While this logic works for Talk and Wikipedia Talk pages, the Hebrew and Malay Wikipedias have more messages on User Talk pages than Japanese. Size is probably not the only explanation of the differences among the various Wikipedias.

4. FINDINGS AND DISCUSSION

We use aggregated data to explicate the similarities and differences between distinct cultures and sizes of Wikipedias. As a part of the process, some codes were grouped into the following categories: quality and accuracy; courtesy; conflict and disagreement. Table 2 presents percentages of codes appeared for specific categories. The percentages were calculated by using code frequencies divided by the total number of posts in order to compensate for size differences among four Wikipedias.

Table 2. West/East and Large/Small comparison on four categories in percentage

Culture—Quality and Accuracy			
	Talk	**User Talk**	**Wikipedia Talk**
West	69.67	45.33	37.72
East	45.24	32.02	29.95
Culture—Courtesy			
	Talk	**User Talk**	**Wikipedia Talk**
West	21.29	120.73	21.54
East	106.87	234.97	114.44
Culture—Conflict and Disagreements			
	Talk	**User Talk**	**Wikipedia Talk**
West	98.82	26	36.22
East	31.39	11.11	6.54
Size—Quality and Accuracy			
	Talk	**User Talk**	**Wikipedia Talk**
Large	52.27	31.72	28.64
Small	62.64	45.63	39.03
Size—Courtesy			
	Talk	**User Talk**	**Wikipedia Talk**
Large	107.45	254.89	113.92
Small	21.33	100.81	22.06
Size			

[1] Due to space limit, the coding scheme was not included in this paper.

	Talk	User Talk	Wikipedia Talk
Large	22.06	10.69	7.34
Small	108.15	26.42	35.42

As can be seen in Table 2, we aggregated the frequencies of codes by culture (East/West) and size (Small/large). Table 3 presents the results of cross tabulation of three namespaces by size and by culture. Western cultures include English and Hebrew Wikipedias and Eastern cultures include Japanese and Malay Wikipedias. The Eastern countries selected here are characterized by high scores in Hofstede's Power Distance Index and high scores in Collectivism. The Western countries, on the other hand, have low scores in Power Distance Index and high scores in Individualism.

In addition, descriptive data indicate that the content of messages varies across types of talk pages and that the pattern of variation is similar across all four languages. For example, postings about accuracy and quality of information are more frequently found on Talk pages than on User Talk pages, or Wikipedia Talk pages. This tendency is evident in both large and small Wikipedias and on Western and Eastern Wikipedias. Communication style, however, is consistent across talk pages in each language but differs based on culture and size between Wikipedias. These variations correlate with Hofstede's dimensions of cultural diversity, specifically with the power distance index.

In terms of similarities across culture and sizes, postings about accuracy and quality of information are more common on Talk pages than User Talk or Wikipedia Talk, in large and small Wikipedias and in Eastern and Western Wikipedias. At the same time, more courteous posts are found in User Talk pages than Talk or Wikipedia Talk pages. We speculate that this pattern reflects the different roles of each type of Talk pages. Talk pages are task oriented as reflected through the emphasis on quality and accuracy while User Talk pages are more social in nature. This pattern is evident across cultures and sizes.

Variations in communication style were found based on Wikipedia size and culture, but at the same time these styles were consistent across all three types of talk pages (Talk, User Talk, and Wikipedia Talk) in each language. Hofstede's dimensions of cultural diversity can explain these variations.

On all three Talk pages, a higher frequency of courtesy behaviors in the large Wikipedias compared to the small Wikipedias was observed (see Table 2); these were found to be statistically significant (Table 3). Variations of courtesy codes across cultures also reveal that the East has significantly more courteous messages on each type of talk pages than the West. Countries differ in the perception of inequalities among its members in the context of family, school, and work. In high power distance countries (i.e., Japan) there would be more respect toward parents (by their children), teachers (by their students), and bosses (by their subordinates). This respect translates to more politeness. Even the language systems in high power distance cultures emphasize distinctions based on a social hierarchy. Politeness is closely related to power distance (e. g., Brown & Levinson, 1987). Moreover, in collectivistic/ high power distance cultures relationship prevails over tasks (Trompenaars & Hampden-Turner, 1998). This may be the reason for more courtesy instances in the East compared with the West. Likewise, the same rational can be used to explain the difference between East and West in the amount of conflict and disagreement. Perhaps this is because in individualistic countries task prevails over relationships (Trompenaars & Hampden-Turner, 1998).

Table 3. Cross tabulation results of 3 talk pages by size and 3 talk pages by culture.

		Pearson $\chi 2$	Cramer's V	P value
East/West	**Quality & Accuracy (N=260, df=2)**	.446	.041	.8
	Courtesy (N=620, df=2)	24.415	.198	000
	Conflict & Disagreement (N=210, df=2)	2.062	.099	.357
Size	**Quality & Accuracy (N=261, df=2)**	.348	.037	.84
	Courtesy (N=620, df=2)	12.413	.141	.002
	Conflict & Disagreement (N=209, df=2)	3.26	.125	.196

5. CONCLUSIONS

This paper identified the Wikipedias as CoPs and examined three Wikipedia talk pages in four languages. It explained similarities and differences among cultures, sizes, and talk pages by the function of namespaces or by Hofstede's dimensions of cultural diversity.

In general, task-oriented postings, such as quality and accuracy, were found more frequently on Talk pages in both Eastern and Western Wikipedias and small and large Wikipedias than on other two namespaces (Wikipedia Talk and User Talk). On the contrary, community well-being postings, such as courtesy, appeared mostly in User Talk pages in both cultures and sizes compared to Talk and Wikipedia Talk pages. The findings indicated that, in any of these languages, Wikipedia users differentiate the use of each type of discussion area by posting

messages of different natures and purposes. When differences between the Eastern and Western Wikipedias are found, these variations can be explained by Hofstede's (1991) dimensions of cultural diversity.

One of the limitations of the study is an assumption that a Wikipedia in a specific language relates to a national culture of a specific country. The languages used in the Wikipedias, however, do not exactly correspond with a specific country. Another limitation common to cross-cultural research is that multiple researchers, speakers of various languages, coded the data. Even though the inter-coder reliability on the English pages among all three coders was high, it is possible that some of the variations across languages are partially due to the differences among the coders. Despite these limitations, this research sheds light on how CoPs operate by analyzing norms of behaviors. In particular, the four Wikipedias that we examined provided exemplars of CoPs that exist in different cultural environments. Future research should expand the number of languages. As there are few studies of cross-cultural analysis about online CoPs, the Wikipedias provided nice test-beds to examine variations of norms of behaviors in different cultures. Future research should look into cross-cultural analysis of CoPs in various cultures as well as size differences.

6. REFERENCES

[1] Anthony, D., Smith, S. W., and Williamson, T. (forthcoming) Reputation and reliability in collective goods: The case of the online encyclopedia Wikipedia, Rationality & Society.

[2] Ardichvili, A., Page, V., & Wentling, T. (2003) Motivation and barriers to participation in online knowledge–sharing communities of practice, Journal of Knowledge Management, 7, 1, 64–77.

[3] Barab, S. A., MaKinster, J. G., & Scheckler, R. (2004). Designing system dualities: Characterizing an online professional development community. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), Designing for virtual communities in the service of learning. Cambridge, UK: Cambridge University Press.

[4] Brown, P & Levinson, S. C. (1987). Politeness: Some universals in language usage. Cambridge: Cambridge University Press.

[5] Bryant, S. L., Forte, A., and Bruckman, A. (2005) Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia, Proceedings in of ACM Group 2005, November 6-9, 2005, Sanibel Island, Florida.

[6] Dubé, L. Bourhis, A. and Jacob, R. (2006) Towards a typology of virtual communities of practice, Interdisciplinary Journal of Information, Knowledge, and Management 1, 69-93.

[7] Hall, E.T. (1976) Beyond culture, Doubleday, Garden City, NY.

[8] Hew, K. F., and Hara, N. (2007) Knowledge sharing in online environments: A qualitative case study. Journal of American Society for Information Science & Technology, 59,14, 2310-2324.

[9] Hofstede, G. H. (1991) Cultures and organizations: Software of the mind, McGraw-Hill, London, UK.

[10] Hofstede, G.H. (2008) Announcing a new version of the Values Survey Module: the VSM 08. Retrieved January 6, 2009 from http://stuwww.uvt.nl/~csmeets/VSM08.html

[11] Lave, J., and Wenger, E. (1991) Situated learning: Legitimate peripheral participation, Cambridge University Press, Cambridge.

[12] Mayers, M. D., and Tan, F. B. (2002) Beyond models of national culture in information systems research, *Journal of Global Information Systems*, 10, 1, 24-32.

[13] Pan, S. L. and Leidner, D. E. (2003) Bridging communities of practice with information technology in pursuit of global knowledge sharing, Journal of Strategic Information Systems, 12, 71-88.

[14] Pfeil, U., Zaphiris, P., and Ang, C. S. (2006) Cultural differences in collaborative authoring of Wikipedia. Journal of Computer-Mediated Communication, 12, 1, article 5. Retrieved November 2, 2009 from http://jcmc.indiana.edu/vol12/issue1/pfeil.html

[15] Riehle, D. (2006). How and why Wikipedia works: an interview with Angela Beesley, Elisabeth Bauer, and Kizu Naoko. Proceedings of the 2006 international symposium on Wikis, Odense, Denmark. DOI= http://doi.acm.org/10.1145/1149453.1149456

[16] Straub, D., Loch, K., Evaristo, R., Karahanna, E., and Srite, M. (2002) Toward a theory-based measurement of culture. Journal of Global Information Management, 10, 1, 13-23.

[17] Stvilia, B., Twidale, M., Smith, L. C., and Gasser, L. (2008) Information quality work organization in Wikipedia, Journal of the American Society for Information Science & Technology, 59(6), 983–1001.

[18] Tapscott, D. and Williams, A. D. (2008) Wikinomics: How mass collaboration changes everything, Portfolio, New York.

[19] Trompenaars, F., and Hampden-Turner. C. (1998) Riding the waves of culture: Understanding cultural diversity in global business, McGraw-Hill, New York.

[20] Wasko, M. M. and Faraj, S. (2005) Why should I share? Examining social capital and knowledge contribution in electronic networks of practice, MIS Quarterly, 29, 1, 35-57.

[21] Wenger, E. (1998) Communities of practice: Learning, meaning, and identity, Cambridge University Press, Cambridge.

[22] Wenger, E., McDermott, R., and Snyder, W. M. (2002) Cultivating communities of practice: A guide to managing knowledge, Harvard Business School Press, Boston, MA.

Curation in the Curriculum: Equipping the Profession to Ensure the Preservation of Information

Ross Harvey
Graduate School of Library and Information Science
Simmons College
ross.harvey@simmons.edu

ABSTRACT
This paper proposes a new area of professional practice based on preservation, required in the LIS profession because of changes in the ways that libraries operate and of changes in education for librarianship, as exemplified by the iSchools paradigm. It notes the significant similarities between analog preservation and digital preservation, and proposes these as the basis for new curriculum for a curatorial stream.

Categories and Subject Descriptors
K.3.2 Computer and Information Science Education – *curriculum*

General Terms
Documentation, Human Factors.

Keywords
Curriculum, Curation, Preservation.

1. INTRODUCTION
This paper is speculative. It posits a question: if (or should it be when?) the iSchools paradigm becomes the dominant paradigm in schools that educate for library and information science (LIS), what happens to the preservation function of libraries, one of the traditional concerns of librarianship? This paper proposes curation as a useful model for considering this question. It therefore relates to the interests of the iConference 2010 in that it reflects on "the core activities of the iSchool community, including … engagement between the iSchools and wider constituencies" and is concerned with two of the conference's areas of interest, "Information management: … technologies of forgetting and remembering" and " Digital libraries: preserving digital information …".

My speculation was prompted by a comment by Daniel Greenstein reported in September 2009, that "The university library of the future will be sparsely staffed, highly decentralized, and have a physical plant consisting of little more than special collections and study areas."[1] This comment contains a concept that has significant implications for the concerns, research activities and curricula of iSchools: the de-emphasizing of library collections and their preservation.

2. THE CHANGING NATURE OF LIBRARIES
Is Daniel Greenstein right? As reported[1] he also suggests that "Within the decade …groups of universities will have shared print and digital repositories where they store books they no longer care to manage. … Under such a system, individual university libraries would no longer have to curate their own archives in order to ensure the long-term viability of old texts". Whether or not we agree with Greenstein's comments, there is no doubt that university libraries, like other kinds of libraries, are restructuring in response to the changing ways in which information is created, managed, and used. One of the many sources that address this realignment is Derek Law's discussion of challenges and changes facing university libraries. One of Law's conclusions is that "one glaring gap remains, the absence of any acceptable definition of trusted repositories"[2]. This point is noted again later in the paper.

Comments such as these suggest that there will be a reduced role for the preservation of physical collections, except in specialized centers[1]. There will also be an increased role for the preservation of digital collections (for which there is ample supporting evidence that is not noted in this paper), but there are at present "glaring gaps" in the infrastructure to accomplish this role[2].

3. THE PRESERVATION FUNCTION AND CHANGING CURRICULUM
The preservation function is one that has traditionally been considered as central to LIS practice. This is indicated in statements such as "The preservation function – the stewardship of the accumulated knowledge base – represents the central obligation of librarianship"[3] and "The archival functions of collecting and preserving are intrinsic parts of the research library's service"[4]. This central role has in the past been acknowledged in the traditional curriculum of LIS schools. However, it should be understood that the centrality of preservation is not universally acknowledged. An emphasis on preservation is still perceived by some librarians as "a step backwards to a world from which automation, new media, management science and those exciting possibilities of the new technology had rescued them"[5].

Curriculum is changing in LIS schools. The increasing reach and influence of the iSchools paradigm was recently characterized as reflecting "the extent to which LIS schools have engaged and embraced technological change" and as "signifying a paradigmatic shift in the educational and disciplinary philosophy of many schools that historically were the providers of library education"[6]. Does the iSchools paradigm (the set of practices that define a scientific discipline) and the curriculum associated with it give preservation a central role? It is neither explicitly noted in, nor excluded from, the general statement of purposes on the iSchools web site that "expertise in all forms of information is required for progress in science, business, education, and culture. This expertise must include understanding of the uses and users of

information, the nature of information itself, as well as information technologies and their applications"[7]. In fact some iSchools pay considerable attention to preservation in teaching and conducting research in digital curation, for example the School of Information and Library Science at the University of North Carolina at Chapel Hill though its DigCCurr project. The iSchools are not alone in not acknowledging a more central role for preservation; the ALISE Research Areas classification scheme[8] pays scant attention to preservation, which is represented mainly in category 91, "Preservation and Archiving", and implied in other categories, perhaps category 100 "Digital Archive Informatics" (whatever that means), and categories relating to special materials ("26. Archival Collections" and "27. Special Collections/Rare Books", for instance).

4. INCREASING ROLE FOR PRESERVATION

If Greenstein, Law and others are correct, there will an increasing role for preservation in the practice of librarianship, although it will (according to Greenstein at least) be concentrated in the hands of a small number of storage facilities rather than, as is the case now, in most libraries. This leads to a consideration of the skills and other requirements for running large repositories housing both digital and analog (non-digital) materials. These skills and other requirements are, I suggest, different from the technology focus represented by the iSchools paradigm These differences support an argument for new thinking about professional roles and, therefore, of the curriculum of LIS schools and iSchools – thinking that is based on the coexistence of the new technologies, which bring considerable potential benefits, and the traditional services, practices and values as represented by preservation. This rethinking could result in the development of two streams in the profession, the first principally concerned with *adopting, implementing and using new technologies to serve the user* (the "understanding of the uses and users of information, the nature of information itself, as well as information technologies and their applications" of the iSchools agenda), and the second a *curatorial stream*, primarily concerned with maintaining the sources of information, rather than with the means to access and exploit it. This latter role will be one of preserving and ensuring the availability of the sources that contain the information. The curatorial role has a long tradition in the library profession, but has been de-emphasized from the latter part of the twentieth century to the present.

But the term *preservation* doesn't quite cut it to describe what is required. It is redolent of old books, pest control, and an obsession with climate control – remnants of "a world from which automation, new media, management science and those exciting possibilities of the new technology had rescued" librarians[5]. (Ratcliffe's comment ignores many of the facts– Which "new" technology has as exciting a history as that of mass deacidification, with its space capsules and explosions?) In the world of digital information the term *preservation* has associations that limit its applicability. We need to redefine the term to free it from its associations with solid objects so that we can accommodate the preservation challenges of digital information.

5. CURATION AND STEWARDSHIP

The terms *curation* and *stewardship* are useful to consider in this context. These terms are, relatively speaking, free from the associations with physical objects (especially printed books) that the term *preservation* has. This freedom allows us to develop new ways of working that focus on both the physical objects that store information (analog preservation) and the information contained in bit streams whose physical location, if they have one at all, is likely to change frequently. It also suggests a useful way of considering the curatorial stream proposed above, which would have as its primary concern maintaining the sources of information, regardless of their form, to ensure their availability, currently and over time.

What, then, is *curation*? My definition is based on the many life-cycle models that have developed to describe the requirements for managing of digital objects over time. These are plentiful. I find the Digital Curation Centre's Curation Lifecycle Model[9] to be the most useful. This considers information in digital forms (in terms of data, digital objects, and databases) and recognizes the centrality of metadata, planning, and collaboration in managing such information over time. Its key "sequential actions" consider the conceptualizing and planning of projects that generate data, through its ingest into a repository and the actions involved with managing, storing and preserving them over time, to the requirements for accessing and using them, and re-using them. This encompassing view is also apparent in other definitions, such as that of the Institute of Museum and Library Science (IMLS): "digital curation (creation, authentication, archiving, preservation, retrieval, and representation of high-quality data for use and reuse over time)"[10].

For data and information sources in digital form, the term *digital curation* is a more inclusive concept than either digital *archiving* or digital *preservation*. It addresses the whole range of processes applied to data over their life-cycle. Digital curation begins before data are created by setting standards for planning data collection that results in "curation-ready" data – data that are in the best possible condition to ensure they can be maintained and used in the future. Digital curation emphasizes adding value to data sets, through things such as additional metadata or annotations, so they can be re-used. Digital curation involves a wide range of stakeholders cutting across disciplinary boundaries; as well as cultural heritage organizations such as libraries, archives and museums, it also involves funding agencies, government bodies, national data centers, institutional repositories and learned societies.

The term *stewardship* is also a contender. In the context of data, the terms *curation* and *stewardship* "both focus on the data but have different views about the nature of data, their life cycles and relations with their environments"[11]. Curation is principally interested in "organizing and overseeing data holdings" and "deals with guidelines and procedures for data ingestion, archive and delivery". Data stewardship is a larger concept, which "provides a large conceptual framework, an overarching process occurring now but attending to the past and taking into account and influencing the future, stretching from data planning to sampling, from data archive to use and reuse"[11]. Both are concerned with a wider view than just preservation considered only as a technical process isolated from services, policies and stakeholders[12].

6. CONCEPTUALIZING PRESERVATION MORE BROADLY

Curation concepts and the curation lifecycle provide a way of conceptualizing preservation more broadly so that it encompasses both analog preservation and digital preservation. Table 1 summarizes the results of a more detailed comparison of the principles and practices of analog preservation and digital preservation carried out as the basis of a presentation to Harvard Library Staff in 2009. Although this table has limitations, such as being a crude content analysis of various public statements about preservation, it assists in developing a broader understanding of curation.

Table 1. Analog and digital preservation principles and practices compared

OBVIOUS SIMILARITIES	
Analog Preservation	**Digital Preservation**
Obsolescence and degradation of artefacts are always with us	Obsolescence and degradation of artefacts are always with us
Ensuring the longevity of artefacts	Protect data
Ensuring the longevity of the information content stored in artefacts	Maintain ongoing access to digital materials despite technological change
Creation of 'preservation-friendly' artefacts	Negotiate with the creators of material to use open, well-supported standard formats for which access tools may remain available; Conceptualize; Create or receive
Redundancy – multiple copies are also a good thing	Provide adequate data backups and create multiple copies; Multiple copies/redundancy
Security and emergency management	Have disaster recovery contingencies in place
Improving storage environment and maintaining it at controlled levels; Prolonging the life of the artefact through preventive action	Provide stable, secure media storage conditions and proper handling
Reformatting (converting the information to a more stable form); Replacing deteriorated artefacts	Copy data to new media well within the expected media life, and check the accuracy of copying
Careful documentation of the condition of the artifact and of procedures and materials used in treatment	Gather sufficient metadata about the material's technical characteristics and requirements to support its preservation and management; Description and representation information; Enhance the metadata
Ongoing policy and procedures review	Monitor the technological environment for signs that formats etc are becoming obsolete; Monitor for evolving solutions; Preservation planning
Protecting artefacts	Maintain adequate data security and protection from viruses, system attack and unauthorized modification of data
Stabilization of artefacts	Limit the range of formats to be managed
Appraisal	Appraise/Select
Collaboration	Work with or seek help from others to develop solutions; Community watch and participation; Interoperability: 'you are not alone'
Keep the original – we keep the original after we reformat it (for example, retain the artefact after digitizing)	Keep the original (bit-stream, analog after digitising)
Encapsulation – we can enclose artefacts in protective material	Encapsulation (digital files – XML wrappers)
All copying introduces change which needs to be accommodated (for example, in reformatting we emphasize checking and validating of the copy)	Constantly check and validate, because all copying of data (such as migration) introduces change
Authenticity – we strive to maintain the authenticity of the artefact (although we acknowledge this isn't always possible) is a good thing	Decode to uncompressed and save as uncompressed (in addition to keeping the original)

This listing of principles and practices suggests significant similarities. For analog preservation, most of the list is encompassed by an emphasis on the artefact – the physical object – and especially on its characteristic of staying reasonably stable over time. This is expressed particularly in the concept of benign neglect: the idea that most artefacts do not deteriorate rapidly if ignored, thus buying time before preservation treatments need to be applied. It is also apparent in practices and procedures such as those that aim to stabilize the artefact, for instance by using stable materials. The integrity of the artefact, its original state if you like, is maintained as far as possible by practices such as limiting intervention in treatment so that its role as an object of material culture is not detracted from. Storing lesser-used materials off-site in an optimally controlled environment is also based on keeping the artefact for as long as possible, with the implication that in doing so its information content will not become unreadable.

For digital preservation, there is an emphasis on the ability to use (and re-use) the digital object that is not apparent in statements about traditional preservation, presumably because in the preservation of an artefact, its information content is considered to be understandable without modification of the artefact. This emphasis is expressed in actions such as retaining old hardware and software to allow access to obsolete media and data, and those in the 'Access, use and reuse' action of the Digital Curation Lifecycle.

My approach has not been very scientific and my conclusions may not withstand too heavy a scrutiny. They need to be tested more

rigorously. One possibility is to apply a framework for thinking about how archival science and techniques translate to the digital environment, articulated by Ken Thibodeau in a 2008 presentation and used here with his kind permission[13]. Thibodeau's framework has four parts:

- *Keep:* apply established archival science or techniques when the knowledge or technique is valid independent of the context in which it is applied
- *Cut:* don't apply established archival science or techniques when the knowledge or technique is not independent of the context in which it is applied
- *Craft:* adapt or modify archival science or techniques that is fundamentally sound, but has not been articulated appropriately for cyberspace
- *Create:* develop new concepts and techniques needed in cyberspace.

What is now required is the application of this framework to the analog preservation principles and practices defined above, testing them to see if and how they need to be modified to be valid also for the principles of digital preservation.

7. CONCLUSION

The take-home message from this speculation can be stated as follows. Changes in the ways that libraries operate suggest a need for the development of a curatorial stream in the LIS profession. The primary concern of this stream is maintaining the sources of information, regardless of their form, to ensure their availability, now and in the future. Changes in education for librarianship are exemplified by the iSchools paradigm, which does not articulate the requirements of this new focus. The significant similarities between analog preservation and digital preservation, combined with new ways of thinking about curation (especially in the digital context) present a strong basis for new curriculum for a curatorial stream to be developed.

8. REFERENCES

[1] Kolowich, S. 2009. Libraries of the future. Inside Higher Ed (September 24, 2009) http://www.insidehighered.com/news/2009/09/24/libraries

[2] Law, D. 2009. Academic digital libraries of the future: an environment scan. New Review of Academic Librarianship 15, 1 (April 2009), 53-67 (p.64).

[3] Commission on Preservation and Access 1990. Annual Report 1989-90, p.1.

[4] Shoemaker, S. (ed.) 1989. Collection Management Issues. Neal-Schuman, p.66.

[5] Ratcliffe, F. W. 1991. In A Reading Guide to the Preservation of Library Collections (ed. G. Kenny). Library Association, p.4.

[6] Bonnici, L. J., Subramaniam, M. M., and Burnett, K. 2009. Everything old is new again: The evolution of library and information science education from LIS to iField. J. Educ. Libr. Inf. Sci. 50, 4 (Fall 2009), 263-274 (p.273).

[7] http://www.ischools.org/site/charter/

[8] http://www.alise.org/mc/page.do?sitePageId=55727&orgId=ali

[9] http://www.dcc.ac.uk/lifecycle-model/

[10] IMLS 21st Century Librarian grants: 5. Programs to Build Institutional Capacity http://www.imls.gov/applicants/grants/pdf/L21_2010.pdf (p.12).

[11] Karasti, H., Baker, K. S., and Halkola, E. 2006. Enriching the notion of data curation in e-science. Computer Supported Cooperative Work 15, 4 (Aug 2006), 321-358. http://portal.acm.org/citation.cfm?id=1166839 (p.352).

[12] Lavoie, B. F., and Dempsey, L. 2004. Thirteen ways of looking at ... digital preservation. D-Lib Magazine 10, 7/8 (July/August 2004) http://www.dlib.org/dlib/july04/lavoie/07lavoie.html.

[13] Thibodeau, K. 2008. Archival Curation in Cyberspace. Presentation to the Mid Atlantic Archives Conference, November 7, 2008.

BROADBAND DEPLOYMENT AS TECHNOLOGICAL INNOVATION: ASSESSING THE NEEDS OF ANCHOR INSTITUTIONS

Charles C. Hinnant
Information Institute
Florida State University
Tallahassee, FL 32306-2100
1-850-645-8967
chinnant@fsu.edu

Charles R. McClure
Information Institute
Florida State University
Tallahassee, FL 32306-2100
1-850-644-8109
cmcclure@lis.fsu.edu

Lauren H. Mandel
Information Institute
Florida State University
Tallahassee, FL 32306-2100
1-850-645-2196
lmandel@fsu.edu

Nicole D. Alemanne
Information Institute
Florida State University
Tallahassee, FL 32306-2100
1-850-645-5683
nalemanne@fsu.edu

ABSTRACT
High-speed broadband facilitates a vast number of beneficial applications such as voice over internet protocol (VoIP), streaming media, gaming, online government and business services, and other interactive services that require high data transmission rates. While high speed broadband is purported to lead ultimately to social and economic development, a coherent proactive national policy regarding the development and use of broadband infrastructure in rural and underserved areas has been slow to appear in the U.S. The FCC's recent mandate to develop a National Broadband Plan and the $7.2 billion in funds specified in the American Recovery and Reinvestment Act of 2009 (ARRA) for broadband deployment and build-out indicates a significant shift in federal government policy to supporting broadband deployment. As the federal funding increases, local community anchor institutions such as public libraries, schools, and medical facilities will be looked to as drivers of successful deployment and adoption of broadband to local communities.

The purpose of this paper is to identify and discuss issues associated with large-scale technological innovations with emphasis on the widespread adoption of high-speed broadband by community anchor institutions. This will include the evaluation of and planning for broadband expansion and implementation. A case study of public libraries in Florida serves to highlight the means used to assess broadband adoption and implementation issues in community anchor institutions. By examining the factors that assist anchor institutions in deploying large-scale broadband projects, this paper also seeks to identify issues and opportunities for iSchools to play a role in assisting anchor institutions with successful deployment of broadband projects.

Categories and Subject Descriptors
C.2.5 [**Computer Systems Organization**]: Local and Wide-Area Networks – *High-speed (e.g., FDDI, fiber channel, ATM).*

General Terms
Measurement, Economics, Legal Aspects.

Keywords
Technological Innovation, Broadband Adoption, ARRA, Needs Assessments, Broadband Planning.

1. INTRODUCTION
In the global information society, the availability of high-speed broadband is an essential ingredient for the ability of individuals and communities to fully take advantage of new information and communication technology (ICT) applications that require high data transmission rates. Consequently, high-speed broadband is a necessary but not totally sufficient prerequisite for such ICT applications as voice over internet protocol (VoIP), streaming media, gaming, and other highly interactive applications. Coupled with more streamlined business and service delivery models, broadband-enabled applications and services are rapidly becoming ubiquitous in institutional service areas such education (distance education), medicine (telehealth and telemedicine), government (e-government, public health, and public safety), and business (e-commerce) [1].

While high-speed broadband supports a vast number of beneficial applications that facilitate social and economic development, a coherent national policy regarding the development and use of a broadband infrastructure has yet to emerge in the U.S. Moreover, there is little agreement regarding what characterizes "high speed

broadband" since its very meaning is tied to the data transmission requirements for more advanced applications at a given time and context. For example, the National Telecommunications and Information Administration (NTIA) defines broadband as an advertised speed of 768 kilobits per second (kbps) downstream and 200 kpbs upstream [2]. This definition is significantly lower than average transmission speeds advertised by many Internet Service Providers (ISPs). A study by the Organization for Economic Co-operation and Development [3] in 2008 cites the average download speed in the U.S. as 9.64 Mbps which drastically exceeds the requirements as outlined by the U.S. FCC [3]. That average ranked the U.S. 19th worldwide, with Japan leading the world with 92.85 Mbps average advertised download speed [3]. In the global Information Society, U.S. connectivity is much slower than other nations' and the FCC's definition of broadband seems inadequate. For this paper, high-speed broadband Internet refers to the average advertised download speed in the U.S. and not the FCC minimum definition that is insufficient for most highly interactive online services [2].

Broadband deployment and adoption is an important component of the Obama administration's policy of investing in core infrastructure areas. The FCC's mandate to develop a National Broadband Plan and the $7.2 billion in funds specified in the American Recovery and Reinvestment Act of 2009 (ARRA) to further this goal are both evidence of this intent. Furthermore, the $200 million set aside for Public Computer Centers (PCC) within NTIA's Broadband Technology Opportunities Program (BTOP) to assist in establishing and supporting local institutions to serve as community broadband anchors for broadband services shows a political intent to employ new broadband infrastructure and services in order to facilitate social and economic development, especially in rural and underserved areas.

The purpose of this paper is to identify and discuss issues associated with large-scale technological innovations with reference to the widespread adoption of high-speed broadband by community anchor institutions (e.g., schools, libraries, medical facilities). This will include the evaluation of, and planning for, broadband expansion and implementation. A case study of public libraries in Florida is used to highlight the means used to assess both broadband adoption and implementation issues. Ultimately, by examining the factors that assist anchor institutions in deploying large-scale broadband projects, this paper also seeks to identify issues and opportunities for iSchools to play a role in assisting anchor institutions in successfully deploying broadband projects.

2. SITUATIONAL CONTEXT OF ANCHOR INSTITUTIONS AND BROADBAND ADOPTION

2.1 Broadband Deployment as Technological Innovation

Since technological innovation is typically defined as "the situationally new development and introduction of knowledge-derived tools, artifacts, and devices by which people extend and interact with their environment," it seems that a better understanding of the innovation process would inform a more complete understanding of broadband deployment in community anchor institutions (anchors) [4].The innovation process for most complex organizations, such as anchors, is heavily influenced by the inter-relationship of social and technical factors within the organizations' internal and external environments, such as the technology itself, the availability of required resources, the fit with the organization's primary task, and an organization's structural arrangements [5]. Furthermore, examinations of advanced online technologies, such as personalization of services, also indicate that the adoption of new online services is often slowed by limitations of technical expertise and budgetary considerations [6].

A study by Tornatsky and Klein [7] indicates that three primary characteristics are repeatedly associated with the adoption of new technologies: relative advantage, ease-of-use, and compatibility. Similarly, the extent to which a particular technology alters existing organizational processes also plays a role in the innovation process. So-called radical innovations usually involve a significant alteration of an organization's processes or outputs, or significantly impact the organization's key stakeholders [8, 9]. Radical innovations that are clear departures from an organization's technological norms generally experience more risks for failure or setbacks than do technological innovations that involve only incremental changes in an organization's existing technological environment. This may be especially true for the adoption of broadband by anchor institutions in rural and otherwise underserved communities since the anchors and their core stakeholders may not possess or, initially have access to, sophisticated ICT infrastructures [5, 10].

The decision for anchor institutions to adopt a technological innovation, such as high-speed broadband, is predicated on the assumption that organizational decision-makers have sufficient awareness of new technologies to understand their potential benefits. The greater the information and knowledge assets that an anchor has at its disposal, the more likely it is to find new technologies to address operational problems, and the more likely it is to understand and implement the technology [11, 12]. Although all technologies require some learning on the part of the staff participating in the adoption, some technologies create a barrier to diffusion by placing more demands on adopters for new knowledge and skills [13]. In addition to the availability of resources, technologies that have greater levels of congruency with key organizational tasks, such as the relationship between high-speed broadband and the mission of many anchors to provide free public Internet and computing access to a wide audience, may be perceived as more useful and therefore, as a more successful adoption of technology [14, 15].

While technology characteristics, intraorganizational issues, and resource issues play important roles in technological innovation for anchor institutions, anchors also exist within a network of multiple government agencies, nonprofit organizations, and private sector organizations [5, 16]. This highlights that broadband innovation is not only a product of the potential demand from anchor institutions, or their immediate clients, but also a function of how the technology is supplied, or pushed, from other institutions that desire the innovation to take place [10, 17]. Therefore, any consideration of broadband adoption by anchors should also consider the extent to which institutions such as the national government subsidizes or regulates the deployment and adoption of high-speed broadband through its telecommunication policies [5, 18].

2.2 Social and Technical Issues of Broadband Adoption in Community Anchor Institutions

In regards to high speed broadband adoption, anchor institutions such as public libraries, schools, and medical facilities are impacted by rapid increases in the need for higher levels of bandwidth as well the availability, or supply, of such infrastructure services. As a growing number of ICT applications require higher amounts of bandwidth, anchors must consider the data transmission speed of their Internet connections since it directly impacts their ability to meet user and staff application needs [5, 19]. In addition, anchors do not usually serve a single user at one time. Rather, they often simultaneously serve many users on a single broadband connection thus further downgrading the data transmission capabilities and, therefore, the applications available to each user. Assuming that such anchors are aware of any potential deficiencies in their ability to adequately serve their clients, this can serve as clear marker of the need, or potential demand, for high speed broadband in community anchors.

Tied closely to both the awareness and motivation to innovate and the characteristics of a specific ICT is the anchor institution's available resources, such as financial resources, the number of available staff, or the knowledge assets that are required to adopt and implement a respective technology. For example, U.S. public libraries and K-12 schools may apply for E-rate discounts under the Universal Service Fund, Schools and Libraries Program, established by the Telecommunications Act of 1996 in order to maintain and expand public Internet access [20]. These discounts may be applied to selected telecommunications, Internet access, and internal connectivity [21]. This funding subsidy is critical for libraries and schools to sustain the provision of free public access Internet to U.S. communities.

In addition to E-Rate funds, the Federal government is currently making funds available to upgrade public computer center capacity through the ARRA, which includes funding for broadband build-out and public computer center upgrades through the BTOP administered by the NTIA and the Broadband Initiatives Program (BIP) administered by the Rural Utilities Service (RUS) [2]. The NTIA-administered BTOP program provides a targeted funding opportunity for anchors, through its emphasis on their importance in deploying and sustaining the adoption of broadband Internet. The BTOP includes set-aside funds for broadband build-out, public computer center capacity expansion, and sustainable broadband adoption education and training programs. The deployment and adoption of high speed broadband is clearly impacted by not only the need of anchor institutions to acquire higher capacity but also the ability of other institutions such as the federal government to push technological innovation through financial subsidies.

3. CASE STUDY OF BROADBAND NEEDS ASSESSMENT

3.1 Assessing the Need for Broadband

Almost all institutional innovation occurs in response to a perceived demand or need [10, 17]. Assessing the need for new technologies is a precursor to determining current, or future, demand on the part of anchor institutions and their broader communities. Therefore, this section provides an overview of a study (funded by a grant from the State Library & Archives of Florida) conducted to assess the need for high-speed broadband in Florida public libraries in order to carry out advanced E-government and emergency management services [22]. Although this project was limited to Florida public libraries, this section provides an overview for other community anchor institutions to conduct similar broadband needs assessments and services.

3.2 Methodological Approaches to the Needs Assessment Study

Research team members employed a multi-method data collection approach to conduct the needs assessment. Data-collection approaches used in this study include:

- Literature review: Review of the literature regarding public library technology and broadband use and deployment;
- Interviews: Interviews with selected public librarians, emergency management officials, and others knowledgeable about the topic to understand existing broadband connections and configurations in Florida public libraries and obtain feedback related to the usefulness of developed maps that indicate public library Internet connectivity;
- Public library case studies: Selected public libraries described and collected data on current broadband connections and infrastructure, workstation connectivity speeds, and network configurations;
- Public library site visits: Onsite review and tests of workstation connectivity speeds and network configurations at selected public libraries;
- Geographic Information System (GIS) analysis of public library telecommunications: Use of GIS software to manage, analyze and map Florida public library broadband data from the Bill & Melinda Gates Foundation Florida public library technology dataset [23] made available from the State Library; and
- Public library national survey data analysis: Analysis of the Public Library Funding and Technology Access Survey [24] related to technology and broadband use and deployment in Florida public libraries.

These methods were selected for their applicability to an exploratory, statewide public library technology needs assessment.

The study team employed a combination of purposeful and cluster sampling for the study's iterative multi-method data collection efforts. The study was exploratory and purposeful, thus limiting the generalizability of the data. The data collection approaches, however, provided detailed and overlapping findings regarding broadband capacity issues in public libraries. By using such an approach, the study team identified and triangulated perspectives on broadband needs for public libraries from both the public library and user populations, thus ensuring reliable and valid data.

3.3 Findings from a Needs Assessment of Florida Public Libraries

The findings present a preliminary picture of Florida public library broadband connectivity and the extent to which Florida public libraries have adequate broadband Internet access to provide public access Internet and computing and a range of other electronic and networked services. Library outlets across the state report insufficient data transmission speeds and the majority of Florida public libraries report that the number of public access workstations is insufficient to meet patron needs some or all of the time. This situation is more pronounced in rural and suburban public libraries.

Connection speeds impact the level of services libraries can offer the public (see Table 1), and in fact, over 75 percent of Florida public libraries report existing connection speeds are insufficient to meet patron and staff demand. Also, most of the librarians who participated in case studies were unaware of the loss of data transmission speeds between their institutional connection and individual workstations at the individual branches. Only Sarasota County public libraries average connectivity speeds over 50 Mbps (75.94 Mbps), the highest for the state (Figure 1). The next highest average speeds are public libraries in Indian River (50 Mbps), Charlotte (45 Mbps), and Leon (33 Mbps) Counties. Without these speeds, public libraries may be able to provide only minimal E-government and emergency management services such as filling out online forms. Furthermore, they will not be able to support advanced applications such as large volume file transfer, digital video streaming, downloading, and sharing, remote education, and building control and maintenance [25].

Situational factors play a critical role in affecting each library's technology access and services. These factors cannot be ignored when considering how best to assist libraries improve network efficiencies and computer equipment. The current cost and speed of broadband for Florida's public library outlets disable many librarians and libraries from adequately serving their communities. These communities turn to their public library outlets for free and publicly available broadband Internet access to participate in today's Information Society. However, slow Internet connectivity speeds, high broadband costs, and situational factors greatly impact libraries' ability to adequately support public access Internet and computing. The needs assessment study discussed here highlights how the skills associated with the iSchools (especially the study of the intersection of people, information, and technology) can play an important role in assisting public anchor institutions by assessing the overall need to innovate by upgrading their broadband capacity and, ultimately, expanding broadband access in the U.S.

Table 1. Comparison of Internet services possible at different speeds [25]

Speed range	Possible services that can be supported
500 kbps – 1 Mbps	Voice over Internet protocol (VoIP), short message service (SMS), basic email, web browsing simple sites, streaming music using caching, low quality and highly compressed video
1 Mbps – 5 Mbps	Web browsing complex sites, email with larger file attachments, remote surveillance, Internet protocol TV-standard definition (IPTV-SD), small and medium size file sharing, ordinary telecommuting, one channel of digital broadcast video, and streaming music
5 Mbps – 10 Mbps	Advanced telecommuting, large size file sharing, multiple channels of IPTV-SD, switched digital video, video on demand SD, broadcast SD video, two to three channels of video streaming, high definition (HD) video downloading, low definition telepresence, gaming, basic medical file sharing and remote diagnosis, remote education, and building control and management
10 Mbps – 100 Mbps	Telemedicine, educational services, broadcast video SD and some HD, IPTV-HD, complex gaming, telecommuting with high quality video, high quality telepresence, HD surveillance, smart building control
100 Mbps – 1 Gbps	HD telemedicine, multiple educational services, full HD broadcast video, full IPTV channels, video on demand HD, immersion gaming, and telecommuting with remote server services
1 Gbps – 10 Gbps	Research applications, uncompressed HD video streaming telepresence, live event digital cinema streaming, telemedicine with remote control of medical instruments, interactive remote visualization and virtual reality, sharing terabyte size datasets, and remote supercomputing

3.4 Considerations for Community Anchor Institutions

While most results from this needs assessment are not generalizable beyond Florida, the needs assessment methodology is transferable to other institutional contexts. The findings, however, are suggestive of issues and topics that are likely to be significant in other states as well. This study shows the value of a broadband needs assessment for evaluating the current situation and for recommending actions to facilitate successful innovation. Anchors across the U.S. can employ some or all of the methods described to evaluate their own network efficiencies and broadband levels. This is crucial for individual anchors to understand more fully the situational context in which they provide Internet access and services, including any successes, deficiencies, and inefficiencies. Such knowledge not only impacts their ability to more effectively adopt and management broadband technology but also impacts their ability to obtain addition funding resources through BTOP, E-Rate, and other government subsidy programs.

Anchor institutions should consider whether an upgrade in broadband capacity actually translates into meaningful broadband data transmission speeds at the level of individual workstations. A key evaluation metric of success will be the number of extant workstations with connection speeds of less than 768 kbps down and 200 kbps up versus the number of workstations that meet or exceed these speeds after adoption of new higher speed broadband connections. Without careful planning and implementation, it is possible that an upgrade in broadband connection still may not meet the FCC requirements for broadband connectivity at the level of the individual workstation. Poorly designed and deployed networks serving too many workstations, wireless routers, and bandwidth-intensive applications can result in the anchor institution still possessing data transmission rates too low to meet the FCC standards for broadband connection speed at the workstation. This is especially true in anchor institutions such as public libraries and public schools where countless users access the network at one time, many of whom rely on the institution to access bandwidth-intensive applications such as file-sharing, Web 2.0 tools, and E-government forms and services. In such instances, additional onsite assessment may be necessary to re-configure the technology or improve the overall telecommunications infrastructure and capacity for data transmission.

Anchor institutions in other states may have access to GIS, statewide school, library, hospital, and other Internet connection data files and can produce maps depicting the Internet connection speeds and costs for in their state. This process may be facilitated by the nationwide broadband data-mapping project currently being implemented by NTIA. In addition to mapping

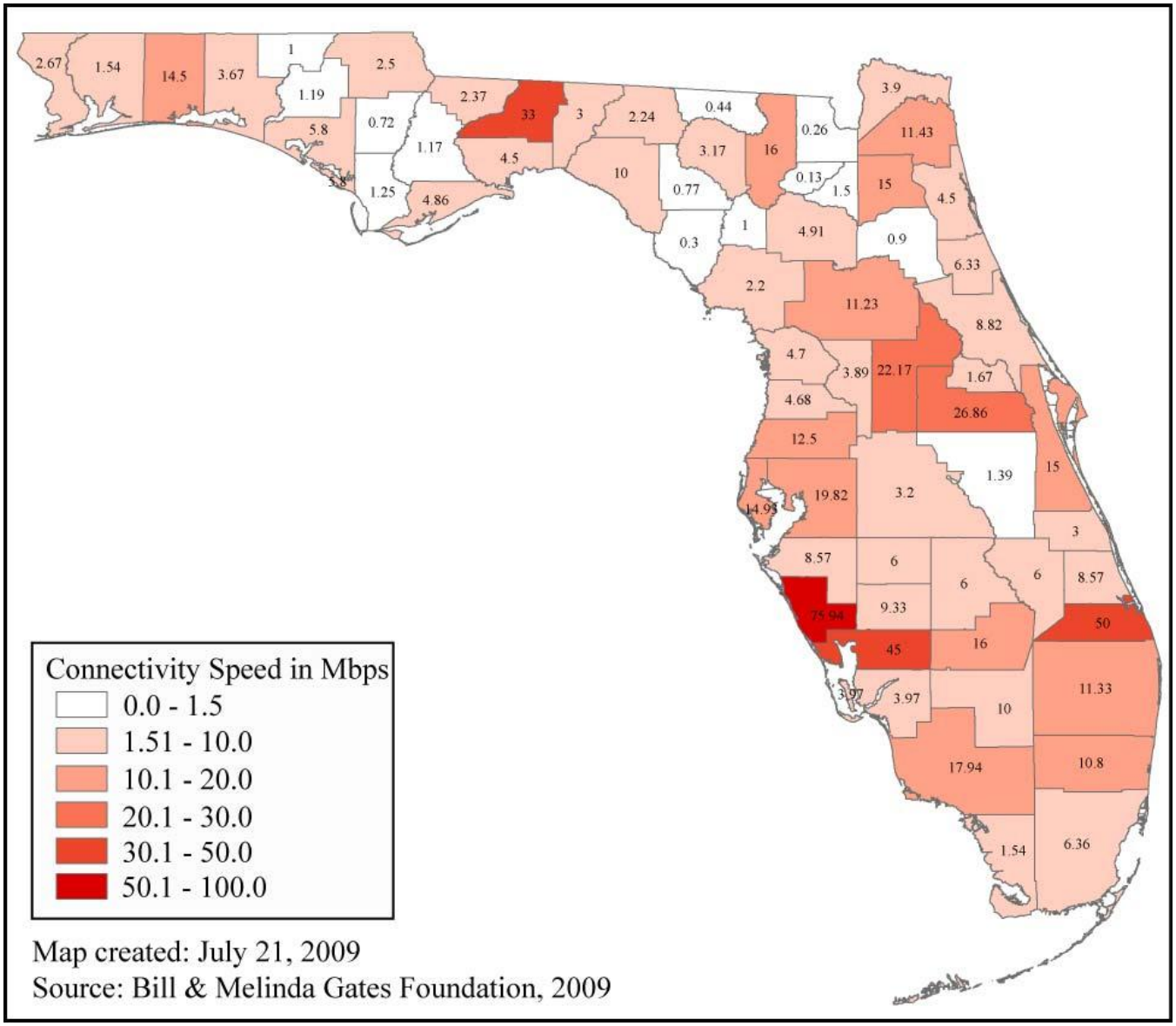


Figure 1. Average connectivity speed for all public library outlets by county: Florida 2009.

Internet connectivity, other anchor institutions may employ the methodology discussed here to better understand the situational factors impacting Internet access and broadband-enabled service provision. Ultimately, this study demonstrates situational factors can be identified, assessed, and planned for prior to the adoption, or upgrade, of high speed broadband. Furthermore, this study shows the necessity of a thorough institutionally-specific assessment of situational factors prior to the adoption of high speed broadband.

4. DIRECTIONS FOR FUTURE RESEARCH

Considering the social and technical factors that influence the technological innovation process, there are several research agendas through which iSchools can assist anchor institutions in successfully adopting and implementing new forms of ICT such as high speed broadband. These can be approached through the following research questions:

- What kinds of planning and evaluation activities should be included in anchors' assessments of needs and planning, as well as the development of funding sources and acquisition strategies?
- How much and what kind(s) of information are necessary for anchors to plan for broadband deployment and adoption?
- By which metrics can anchors assess the level of fit between broadband and the institutional goals it is attempting to address through high speed broadband adoption, and which of these metrics are most appropriate?
- Which resources (e.g., funding, staffing, and government support) do anchors need in order to sustain higher levels of broadband service, and how much of these resources are necessary to sustain different levels of broadband service?
- How do federal, state, and local information policies and regulations affect the success with which anchors can access and deploy broadband?
- At what point is the return on investment of in-house knowledge for broadband deployment low enough to make external expertise the more efficient investment of anchors' resources?

The preceding list offers questions intended to begin the discussion of such agendas. An understanding of the issues associated with technology innovation serves as a necessary foundation from which iSchools can investigate these questions.

5. CONCLUSION

The deployment of high-speed broadband to underserved and rural communities is a crucial technological innovation if such communities are to achieve the economic and social benefits of the so-called Information Society. Without the successful adoption and implementation of high speed broadband, many communities will be unable to make use of services such as VoIP, streaming media, and interactive applications that are almost ubiquitous in many locations across the globe. As the federal government looks to community anchor institutions such as public libraries, schools, and medical facilities to serve as local foundations for high speed broadband deployment, it is imperative to understand better the social and technical factors that will impact the successful adoption of such technologies.

This paper provides a review of selected literature in technological innovation as a means of highlighting how technology characteristics, intra-organizational issues, and the broader innovation environment impact the ability of anchor institutions to successfully adopt and implement high-speed broadband. Furthermore, it has described a case study of public libraries in Florida that illustrates how to assess the social and technological need for high-speed broadband at the level of anchor institutions. While the findings of this case study are not generalizable to other states they are suggestive of issues likely to be in place in other states. The methodology it employs can be a useful guide for anchor institutions in other locations. Such studies are especially crucial as anchors seek to assess and understand the situational contexts in which they provide Internet access and services prior to seeking funding through BTOP, E-Rate, and other government programs. Finally, the discussion presented in this paper highlights the fact that the interests, skills, and expertise possessed by iSchools make them well positioned to play a role in assisting anchor institutions such as public libraries, schools, and medical facilities increase their ability to successfully adopt new broadband technologies as well as the applications and practices they enable.

6. REFERENCES

[1] Communications Commission. n.d. Broadband.gov. Beta. http://www.broadband.gov/.

[2] Broadband Initiatives Program, Broadband Technology Opportunities Program notice of funding availability, 74 Fed. Reg. 33104. 2009. http://www.broadbandusa.gov/files/BB%20NOFA%20FINAL%2007092009.pdf.

[3] Organisation for Economic Co-operation and Development (OECD). n.d. OECD broadband portal [Web portal]. Organisation for Economic Co-operation and Development. http://www.oecd.org/sti/ict/broadband.

[4] Tornatzky, L. G. and Fleischer, M. 1990. The Process of Technological Innovation. Lexington Books.

[5] Papacharissi, Z. and Zaks, A. 2006. Is broadband the future? An analysis of broadband technology potential and diffusion. Telecommunications Policy 30, 1 (Feb. 2006), 64–75.

[6] Hinnant, C. C. and O'Looney, J. 2003. Examining pre-adoption interest in online innovations: an exploratory study of E-service personalization in the public sector. IEEE Transactions on Engineering Management 50, 4 (Nov. 2003), 436-447.

[7] Tornatzky, L. G. and Klein, K. G. 1982. Innovation characteristics and innovation adoption-implementation: a meta-analysis of findings. IEEE Transactions on Engineering Management 29, 1 (Feb. 1982), 28-45.

[8] Dewar, R. D. and Dutton, J. E. 1986. The adoption of radical and incremental innovations: an empirical analysis. Management Science, 32, 11 (Nov. 1986), 1422-1433.

[9] Ettlie, J. E., Bridges, W. P., and O'Keefe, R. D. 1984. Organizational strategy and structural differences for radical versus incremental innovation. Management Science 30, 6 (Jun. 1984), 682-695.

[10] Frieden, R. 2005. Lessons from broadband development in Canada, Japan, Korea and the United States. Telecommunications Policy 29. 8 (Sep. 2005), 595–613.

[11] Nilakanta, S., and Scamell, R. W. 1990. The effect of information sources and communication channels on the diffusion of an innovation in a data base environment. Management Science 36, 1 (Jan. 1990), 24-40.

[12] Fichman, R. G. and Kemerer, C. F. 1997. The assimilation of software process innovations: an organizational learning perspective. Management Science 43, 10 (Oct. 1997), 1345-1363.

[13] Attewell, P. 1992. Technology diffusion and organizational learning: the case of business computing. *Organization Science*, 3, 1(Feb. 1992), 1-19.

[14] Cooper, R. B. and Zmud, R. W. 1990.). Information technology implementation research: a technological diffusion approach. Management Science 36, 2 (Feb. 1990), 123-139.

[15] Falch, M. and Tadayoni, R. 2007. Next generation broadband – content and user perspectives. Telematics and Informatics 24, 4 (Nov. 2007), 243–245.

[16] Picot, A. and Wernick, C. 2007. The role of government in broadband access. Telecommunications Policy 31, 10-11 (Nov.-Dec. 2007), 660–674.

[17] King, J.L., Gurbaxani, V., Kraemer, K.L., McKarlan, F.W., Raman, K.S., and Yap, C.S. 1994. Institutional factors in information technology innovation. Information Systems Research. 5, 2 (Jun. 1994), 139-169

[18] Cambini, C. and Jiang, Y. 2009. Broadband investment and regulation: a literature review. Telecommunications Policy 33, 10-11 (Nov.-Dec. 2009), 559–574.

[19] Bertot, J. C. and McClure, C. R. 2007. Assessing sufficiency and quality of bandwidth for public libraries. Information Technology and Libraries 26, 1(Mar. 2007), 14-22.

[20] Telecommunications Act of 1996. 1996. 110 Stat. 56 § 706. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=104_cong_bills&docid=f:s652enr.txt.pdf.

[21] McClure, C. R. and Jaeger, P. T. 2009. Public libraries and Internet service roles: measuring and maximizing Internet services. American Library Association.

[22] McClure, C. R., Mandel, L. H., Snead, J. T., Bishop, B. W., and Ryan, J. 2009. Needs assessment of Florida public library E-government and emergency/disaster management broadband-enabled services. Florida State University, College of Information, Information Use Management & Policy Institute. http://ii.fsu.edu/content/view/full/14970.

[23] Bill and Melinda Gates Foundation. 2009. Florida public library technology dataset [Electronic database]. Bill and Melinda Gates Foundation.

[24] Bertot, J. C., McClure, C. R., Wright, C. B., Jensen, E., and Thomas, S. 2009. Public libraries and the Internet 2009: study results and findings. Florida State university, College of Communication and Information, Information Use Management & Policy Institute. http://ii.fsu.edu/content/view/full/17025.

[25] California Broadband Task Force. 2008. The State of Connectivity: Building Innovation Through Broadband. California Broadband Initiative. http://www.calink.ca.gov/pdf/CBTF_FINAL_Report.pdf.

IntegraL: An Effective Link-based Federated Search Infrastructure

Shuyuan Mary Ho, Min Song, Michael Bieber, Eric Koppel, Vahid Hamidullah, Pawel Bokota
New Jersey Institute of Technology
College of Computing Sciences, Department of Information Systems
University Heights, Newark, NJ 07102
(973) 596-3000
[smho, song, bieber, erk7, vh22, pmb9] @njit.edu
http://smho.mysite.syr.edu http://web.njit.edu/~bieber http://web.njit.edu/~song

ABSTRACT
This research provides a new means for making digital library services interoperable. IntegraL facilitates a virtual restructuring of public web spaces and services, bringing authenticated digital libraries into broad "federated" digital library spaces constructed from numerous interrelationships. Elements of users' search interests reside within a rich context of meta-information that helps users understand and work with them. This provides a ripe environment for organizations and individual people to develop small, specialized collections and services, which automatically become part of the federated space and accessible to those they can benefit. IntegraL extends the boundaries of how we think about and interact with digital libraries.

General Terms
Systems, Design, Experiments

Keywords
Digital library infrastructure, Federated search

1. INTRODUCTION
IntegraL is a scalable, "lightweight" search infrastructure that brings a plethora of relevant resources directly to library users. IntegraL virtually integrates collections and services, including search services of libraries nationwide. It helps users to effectively search structured content information based on identified name entities across heterogeneous digital libraries. Users not only interact with the libraries and search engines just as before, but also see extra link anchors. Upon selecting one link anchor, IntegraL automatically generates a list of links to relevant documents, services and metadata (Song & Bieber 2008). IntegraL further provides recommendation features for search users (Im & Hars 2007). IntegraL allows the library systems to act as *information requesters* (a customized set of links embedded in display screens that widens search horizontally) and *information providers* (link anchors leading to a systems' documents and services vertically). IntegraL provides federated search across all relevant resources and helps users locate information more effectively. It also increases the accessibility and effective usage of library resources.

This paper contains six major sections describing this study. In the following section, we review why an infrastructure is necessary for integrating multiple digital libraries. In the third section, we describe the IntegraL system architecture and IntegraL's innovative way of expanding the use of multiple digital libraries, databases, and search engines at the user's preference. We present our research questions and hypotheses in the fourth section. In the fifth section, we describe the design and execution of a user study for this virtual integration infrastructure. In the sixth section, we discuss the results of our hypotheses testing, and conclude our study in the seventh section.

2. RELATED WORK
Rao (2004) described the progression of information search from the 60's to the 90's. Users' information search has been drastically enhanced from simple query-in, result-out in the 60's, to information digest, indexing, extraction, categorization, visualization, and further to federated research. While users' search capability has been empowered, the design and development of digital libraries have become more sophisticated. Information retrieval will be more based on open and flexible infrastructures (Kazai & Doucet 2008; Rao 2004). However, this scalable archival infrastructure facilitates the collaboration among heterogeneous digital libraries. Being able to accurately retrieve documents from distributed uncooperative digital libraries becomes critical with foreseen and unforeseen problems. They include issues with archival preservation of digital content, indexing in each collection, representable query phrase, merging and transforming retrieved data, effective use of metadata for search, robust retrieval algorithms, seamlessness interactions between user and the data, integration between services and tools, and inevitably privacy and security considerations when accessing data for sensitive purpose.

Merging results from different databases and search engines requires acquisition of database resource description, selecting from collection, and merging results into a single rank list (Si & Callan 2002). In merging high volume data streams in a web-based infrastructure, Mazzucco and Ananthanarayan (2002) uses data mining to processing streams of data, extract patterns and anomalies.

Federated search (or, distributed information retrieval) has been discussed extensively in the research community. Open Archival Initiative (OAI) is a framework that provides search services over aggregated metadata among federated digital libraries for both service providers and data providers (Lagoze & Van de Sompel

2001). Maly and Zubair, et al. (2005) researched a Grid-based architecture for parallel harvesting among large amounts of computing resources to be shared across organizational boundaries. This federated digital library architecture indexes and harvests hundreds of metadata from data providers. This type of architecture requires extensive load balance and may still suffer insufficient service performance if low bandwidth of the data providers and high volume of the harvest nodes are encountered. On the other hand, a user modeling approach to full-text federated search focuses on collecting user behavior by analyzing a user's long-term persistent interests based on user's past queries (Lu & Callan 2006). While this approach may enhance the robustness of the search, redundant documents from the static search collections may lead to unnecessary processing costs in federated search in an uncooperative environment. Shokouhi and Zobel (2007) studied how different queries used can reduce the overlap in search results from dynamic collections. Shokouhi, Baillie, et al. (2007) further studies accurate retrieval in updating dynamic search collection for federated search. Different retrieval process algorithms that enhance recall and precision are studied in uncooperative distributed search environments (Callan & Connel 1999; Callan, Lu et al. 1995; Paltoglou, Salampasis et al. 2007; Paltoglou, Salampasis et al. 2008; Si & Callan 2003; Si & Callan 2005).

Not only are the synchronous operations of the architecture among multiple harvesting nodes important to federated search, the ability to harvest item-level metadata would also enhance the performance of federated search. The advent of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) helps the sharing and harvesting of item-level descriptive metadata for selected digital resources (Arms, Dushay et al. 2003; Hagedorn 2003; Simon & Bird 2003). Foulonneau and Cole, et al. (2005) states that this granular collection-level descriptive metadata provides attributes to the retrieved documents, which would bring more relevant documents in response to a query.

Wu and Li, et al. (2006) suggested noun phrase used as key-phrase in automatic text extraction; these key-phrases can be used as document metadata for web searching (Li, Wu et al. 2004; Wu, Li et al. 2006). Bot and Wu, et al. (2005) used these extracted key-phrases as topical oriented categories from the retrieved documents to correspond to different semantic aspects of the query. Highlight, as one example of metadata search engine, is composed of document acquisition, document pre-classification and automatic concept hierarchy generation. Document acquisition retrieves documents in response to a query. The document pre-classification module classifies documents into pre-defined categories. Then, based on the extracted key-phrases, the hierarchy module automatically generates individual concept hierarchies within each active category (Bot, Wu et al. 2005).

3. SYSTEM ARCHITECTURE

IntegraL offers a scalable infrastructure that links among heterogeneous digital libraries through the development of a web-based proxy server, sitting on top of Tomcat, an open source servlet container as the outer dotted line represented in Figure 1. When a user makes a request to access a digital library, he or she is authenticated by a single sign-on (SSO) mechanism. We adopt an open source authentication mechanism, Shibboleth, in order to allow seamless browsing across digital libraries that have also adopted Shibboleth. Once the user is authenticated, this SSO mechanism allows the user to surf among various subscribed digital libraries without going through repetitious logins at each stage. The user's credential information is stored in hash files on the proxy.

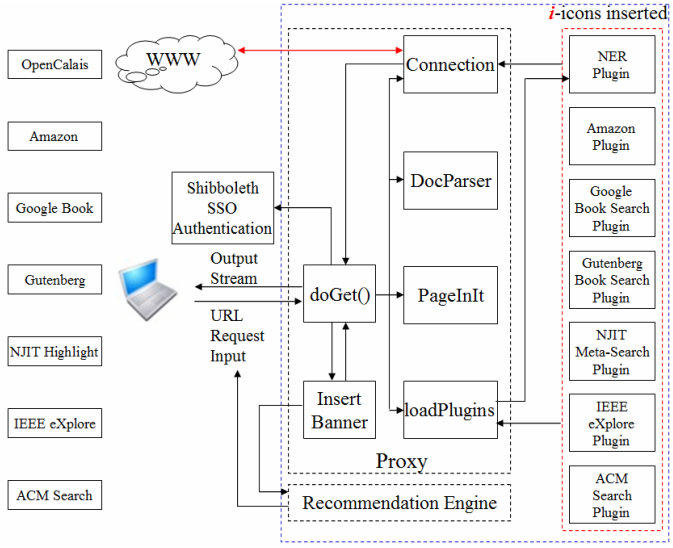


Figure 1: System Architecture

All users' requests are handled by the proxy. When a user browses any other subscribed digital library, their request to access the other digital library is taken care of by the proxy. If a user requests to use other search engines, the pages of their request are returned to the user untouched. When a user makes a URL request, an IntegraL banner is inserted on top of the digital library. All HTML pages are converted into DOM Documents; all relative URL paths are converted to absolute URL paths. What we innovatively create is the addition of i-icons (Figure 2). The i-icons are inserted whenever a name entity is recognized by OpenCalais, a service that annotates data with rich semantic metadata. The name entity recognition module, or NER plugin, receives the document requested by the user from the proxy, analyzes the lexical meaning of recognized categories, such as person, location, etc., and then creates rich semantic metadata that serves as recommended search for user's further references.

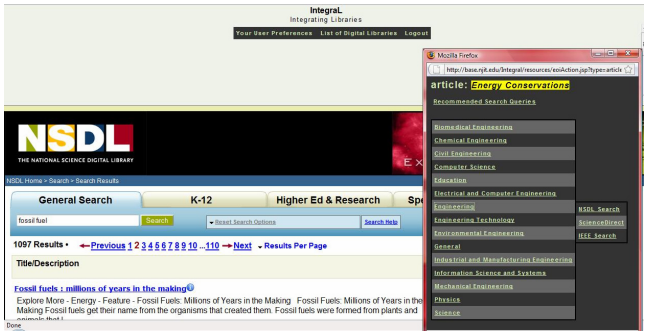


Figure 2: IntegraL links with NSDL

Illustrated on the right-hand side of the Figure 1, plugins are developed using XPath, which parse the HTML documents and insert i-icons wherever elements of interests have been located. The development of plugins is based on templates. This provides flexible expansion of integrating search engines and digital

libraries. IntegraL's web-based administrative interface empowers users with ad hoc configuration.

The strengths of IntegraL lie in its capability to include a federated search engine and recommendation engine that empowers users' deep search. The elements of interests (EOIs) allow ease of integration among heterogeneous databases and library resources. The function of name entity recognition plugins are dynamically plugged in to recognize more named entities: (1) EOIs (elements of interests) are statically defined by the layout-based plugins in the proxy. The plugin templates allow ease of integration among heterogeneous databases. (2) The NER (name entity recognition) plugins are dynamically plugged in to recognize more named entities. (3) Its capability to include the federated search engine and recommendation engine that empowers users' deep search.

4. RESEARCH QUESTIONS

In order to understand how effectively IntegraL can assist users to conduct advanced, deep search through this virtual integration of libraries, we plan to answer the following research questions.

1. Can enhanced access to information through virtual integration help users find information more effectively and to perform tasks involving library resources more effectively?
2. Can enhanced access to information through virtual integration increase the accessibility and utilization of library resources?

4.1 Hypotheses Testing

We form three research hypotheses in order to answer our research questions.

H1: Users perform better objectively when they use IntegraL than without IntegraL.

H2: Users perceive more effective searching subjectively when they use IntegraL than without IntegraL

H3: Users utilize and access more various virtually integrated databases with IntegraL than without IntegraL.

5. METHOD

5.1 Experiment Design

A *within-subject* experiment using a step-by-step approach that helps users learn how to search scholarly articles.

5.2 Data Collection

We conducted a pilot usability study as well as a main experiment during Fall 2009. The total of 139 participants from freshmen in a Physics class, above 18 years of age, participated in the search experiment in October 2009. 7-point Likert scale survey instruments were used. All data were collected in three stages: before, during, and after the experiment. Four different ways were designed to collect our data.

Four different were designed to log our data.

1. Participants' demographics were surveyed.
2. Participants' clickstreams were logged.
3. Participants' objective performance that measured the quality of their search, including relevancy of reference and citation, the use of scholarly database, participants' own judgment of the relevancy of the articles found, and participants' own reasons for ranking the reference lists. The performance measures were rated by 4 graders with Cronbach's alpha value of 0.965.
4. Participants' perceived satisfaction and effectiveness of their search were collected in the post-task survey.

5.3 User Task

All participants were given two tasks to complete. The two tasks were designed with similar search steps. Each participant took about 20-25 minutes to work on each task. The amount of time for the completion of these two tasks was totaled as 45-50 minutes maximum. We scheduled a five minute orientation session before the experiment started; this helped participants get situated easily.

5.4 Rotation of Conditions and Tasks

There are altogether 8 combinations of two tasks (T1, T2) rotated within one treated condition with IntegraL, and one baseline condition for all participants. In order to evenly distribute task assignments with different combinations of systems conditions, the system conditions are controlled centrally at the proxy server and the rotation of the task assignments are combined and controlled at the handouts. In other words, there are eight combined system conditions and there are four sets of task assignment combinations. Table 1 illustrates how these combinations of task assignments and system conditions should be randomized. With this rotation design of user's tasks and system conditions, the threats to validity occurring within subjects can be eliminated.

Table 1: Rotation of Conditions and Tasks

Tasks Rotation			Systems Conditions	
T1	T2	×	IntegraL	Baseline
			Baseline	IntegraL
T2	T1		IntegraL	Baseline
			Baseline	IntegraL
2 sets of tasks rotation			4 sets of system conditions	
2 (tasks) * 4 (conditions) = 8 combined tasks & conditions				

5.5 Data Analysis

The data we used to test the hypotheses were from the experiment in Fall 2009. Data were aggregated and cleaned based on the criteria whether all four aspects of data were completely collected. Data were synchronized and reorganized based on a unique identifier. A complete and useful dataset was reduced to 65 out of the 139 dataset. This small sample size contained confounding effects and affected our hypotheses testing results. Descriptive statistics were conducted to understand general participants' demographics, user's perceptions, and objective measures. We used open source R statistics tool and SPSS to run our analysis. Hypotheses were tested. The results showed that null hypotheses were rejected, and our research hypotheses were supported.

5.6 Threats to Validity

This experiment does contain some threats to validity. However, we were able to justify in our research design how to reduce those foreseeable threats.

1. The design of systematically rotating tasks and conditions eliminates threats to internal validity.

2. The grading rubric provides initial guiding principles of judging user search performance. This inter-rater reliability eliminates threats to criterion validity regarding a user's objective search quality and search performance.
3. Participants usually have high skills in experiencing multiple search engines. This pre-existing search engine experience creates inequality judgments and participants' bias when users come to use a new different ways of search. This threat was taken care of in the design of how we answer our research question. We choose to use objective outcome measures on users' quality of search rather than using objective clickstream and the amount of time spent on completing one task, because the clickstream and the amount of time spent on using IntegraL depict users exploration of this virtual integration of library services, rather than completing a task.

6. DISCUSSIONS

6.1 Existing Library Experience

On a 7-point Likert scale, with 1 as least experienced and 7 as most experienced, most of 65 participants were very skillful with the mean of 5.94 in using the search engines (Figure 3). However their experience of using a digital library such as NSDL, ACM, IEEE and Science Direct, was very slim with the mean 3.45 (Figure 4).

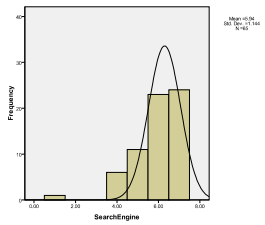


Figure 3: Existing Search Engine Experience

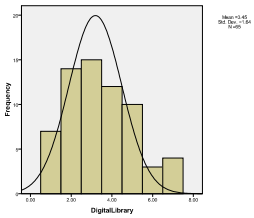


Figure 4: Existing Digital Library Experience

6.2 Outcome Measures of User's Performance

Users generally perform about the same level of search results, however if we compare user objective performance in quality search (Figure 5), users who used IntegraL virtual services would perform higher quality search results than without.

The positive skew distribution presented in Figure 6 depicts that IntegraL helps to enhance search productivity. Users find IntegraL useful in locating relevant information. IntegraL also helps users to do tasks more effectively.

Participants' performance were measured by 4 graders with Mean of bivariate correlations R value, 0.839 (p=0). Because of the *within-subject* experimental design, each participant worked on one task in a treatment condition (with experience of IntegraL), and the other task in the baseline condition (without experience of IntegraL). When comparing the outcome performance measures between the treatment condition and baseline condition, we discovered that participants who used IntegraL system are likely to obtain higher performance measures than those participants who did not use IntegraL system, $t(128) = 1.409$, p=0.002. The t value of 1.409 is not more extreme than the cutoff t of 2.364. Our findings could not reject null hypotheses and this is probably caused by our small sample size of only 65 participants in dataset.

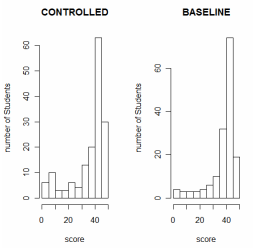


Figure 5: Comparison of User's Objective Performance

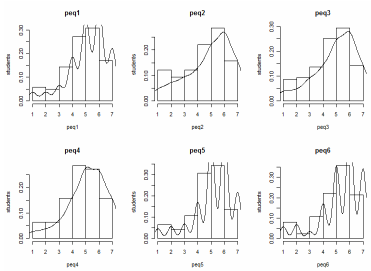


Figure 6: Perceived Effectiveness of Search

However, we conducted a one-way ANOVA estimating population variance in the outcome performance measures from variation within each sample. The $F(1, 128)=10.097$, p=0.002 is more extreme than the cutoff F of 6.64, meaning we could reject the null hypothesis; the research hypothesis that users perform better objectively when they use IntegraL than without IntegraL is supported.

H1: *Users perform better objectively when they use IntegraL than without IntegraL.*

6.3 Users' Attitude and Perceived Acceptance

Users have neutral attitude toward IntegraL, which is possibly due to many good search engines available to users, and that users tend to be used to existing ways of search (Figure 7).

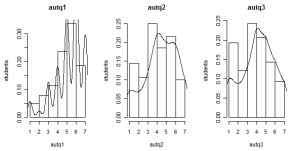


Figure 7: Attitude

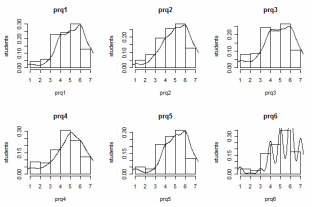


Figure 8: Perceive Performance of IntegraL over other search engines

Users are confident of IntegraL's ability to provide satisfactory search results as illustrated in Figure 8. When comparing IntegraL with other regular search engines as designed in our baseline condition, most users perceive IntegraL to provide better results than those currently available on the market.

We conducted a t-test for a single sample of all independent variables on user perception after they experienced IntegraL system. The results prove all t values are more extreme than the cutoff t of 2.654. Therefore, reject the null hypothesis; the research hypothesis is supported.

We then compared user's subjective perception from before the task to after the task. We conducted a t-test for paired samples, and derived $t(64)=1.769$, p=0.082. The t value 1.769 is more extreme than the cutoff t of 1.669 (two-tailed t-test). Therefore, we reject the null hypothesis; the research hypothesis is supported.

H2: *Users perceive effective searching subjectively when they use IntegraL than without IntegraL.*

We conducted a principal factor analysis to reduce construct dimensions and identified 5 factors of which the Eigenvalues are above 1 (Table 2).

Table 2: Principal Factorial Analysis

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	14.363	59.848	59.848	14.363	59.848	59.848	4.807	20.029	20.029
2	1.780	7.419	67.266	1.780	7.419	67.266	4.796	19.982	40.011
3	1.331	5.547	72.813	1.331	5.547	72.813	4.670	19.459	59.470
4	1.122	4.675	77.488	1.122	4.675	77.488	3.325	13.856	73.326
5	1.026	4.274	81.762	1.026	4.274	81.762	2.025	8.436	81.762

Extraction Method: Principal Component Analysis.

Then, we used Varimax with Kaiser Normalization to run a factor analysis of a correlation matrix and identified the survey response items to be closely correlated. All the items were positively correlated with statistical significance.

6.4 Users' Effective Utilization of Virtual Integration

In order to find out whether users access more various virtually integrated resources, we compared the number of clicks for baseline and treatment conditions, and the amount of time spent for baseline and treatment conditions. We conducted a two-tailed t-test for independent means for two groups on the number of clicks, $t(128)=-1.703$, $p=0.091$. The t value -1.703 is more extreme than the cutoff t of -1.645. Therefore, we reject the null hypothesis; the research hypothesis is supported. We also ran a two-tailed t-test for independent means for two groups on the amount of time spent, $t(128)=-4.855$, p=0.000. The t value -4.855 is more extreme than the cutoff t of -2.576. Therefore, we reject the null hypothesis; the research hypothesis is supported.

We conducted a one-way ANOVA estimating population variance in the number of clicks from variation within each sample. The $F(1, 128)=0.046$, $p=0.831$ is not more extreme than the cutoff F, we could not reject the null hypothesis; the research hypothesis that users access more various virtually integrated databases than without IntegraL is not supported. This finding showed that if we use only the number of clicks to measure the amount of accessibility to other library resources through IntegraL virtual integration is still insufficient. We may need to combine hypothesis 3 with hypothesis 2 (which involves more participants' subjective views.) But, we conducted a one-way ANOVA estimating population variance in the amount of time spent from variation within each sample. The $F(1, 128)=6.568$, $p=0.012$ is more extreme than the cutoff F, we could reject the null hypothesis; the research hypothesis is supported that users tend to utilize more various virtually integrated databases than without IntegraL.

H3: *Users utilize and access more various virtually integrated databases with IntegraL than without IntegraL.*

Moreover, participants of this experiment gave us feedback about their experience of this infrastructure that provides the virtual integration among digital resources. Below are some quotes from the participants.

> "*I really enjoyed using IntegraL because other search engines I have used never gave me relevant references I was looking for. It takes longer to search for your topic using Google, Ask.com, etc whereas IntegraL makes it easier, faster, and it really beneficial.*"
>
> "*I liked using this method because it is easier to find similar topics.*"
>
> "*IntegraL was easier to use than that of Google or yahoo. I was happy with the I-icon and the availability of other resources at one time.*"
>
> "*The Integral system is very useful. It compiles many different search engines in one site so that you do not have to go about looking separately for them. They also give articles from credited sources so that you know they are professional journal entries.*"
>
> "*IntegraL was a very helpful system, it definitely cut my searching time in half and it allowed me to find precise articles pertaining to my article.*"

7. CONCLUSION

Without the complication of complete system integration, IntegraL adopts a light-weight approach that links multiple heterogeneous digital libraries and search engines. It allows interoperability among different search results, search engines and digital libraries. The system itself is mostly built on open-source software, which can be reliable, auditable and cost-effective. This approach provides recommendations to users for further deep search based on identified elements of interests among wide ranges of virtual resources.

Our study showed that participants *perceive more effective searching subjectively when they use IntegraL than without IntegraL* (hypothesis 2), this answered our first research question that enhanced access to information through IntegraL virtual integration would help users find information more effectively. Our study also showed that participants *perform better objectively when they use IntegraL than without IntegraL* (hypothesis 1). This also demonstrates that IntegraL helps users to perform tasks involving library resources more effectively. We also answered our second research question that enhanced access to information through IntegraL virtual integration has helped the users to *utilize more various virtually integrated databases than without IntegraL* (hypothesis 3) and users *perceive IntegraL to be effective in searching activities* (hypothesis 2).

8. ACKNOWLEDGEMENTS

Partial support for this research was provided by the National Science Digital Library (NSDL) under grant LG-02-04-0002, by the National Science Foundation under grants DUE-0434581 and DUE-0434998, and by the New Jersey Institute of Technology. The authors thank Xiangmin Zhang for his advice on the

experimental design, and many graduate assistants who worked on data collection and analysis of this project.

9. REFERENCES

[1] Arms, W. Y., Dushay, N., D., F. and Lagoze, C. (2003) A case study in metadata harvesting: the NSDL., Library High Tech, 21, 228-237.

[2] Bot, R. S., Wu, Y. B., Chen, X. and Li, Q. (2005) Generating Better Concept Hierarchies Using Automatic Document Classification, CIKM'05, Bremen, Germany, pp. 281-282.

[3] Callan, J. P. and Connel, M. (1999) Query-based sampling of text databases, ACM Transaction of Information Systems, 19, 97-130.

[4] Callan, J. P., Lu, Z. and Croft, W. B. (1995) Searching distributed collections with inference networks., SIGIR'95, New York, NY, USA, pp. 21-28.

[5] Foulonneau, M., Cole, T. W., Habing, T. G. and Shreeves, S. L. (2005) Using collection descriptions to enhance an aggregation of harvested item-level metadata, JCDL'05, Denver, Colorado, pp. 32-41.

[6] Hagedorn, K. (2003) OAIster: a "no dead ends" OAI service provider, Library High Tech, 21, 170-181.

[7] Im, I. and Hars, A. (2007) Does a one-size recommendation system fit all? the effectiveness of collaborative filtering based recommendation systems across different domains and search modes, ACM Transactions on Information Systems (TOIS), 26.

[8] Kazai, G. and Doucet, A. (2008) Overview of the INEX 2007 Book Search track: BookSearch '07.

[9] Lagoze, C. and Van de Sompel, H. (2001) The Open Archives Initiative: Building a low-barrier interoperability framework, Proceedings of teh First ACM/IEEE Joint Conferent on Digital Libraries, Roanoke, VA.

[10] Li, Q., Wu, Y. B., Bot, R. S. and Chen, X. (2004) Incorporating Document Keyphrases in Search Results, The 10th Americas Conference on Information Systems, New York, New York, pp. 1-8.

[11] Lu, J. and Callan, J. P. (2006) User Modeling for Full-Text Federated Search in Peer-to-Peer Networks, SIGIR'06, Seattle, Washington, pp. 332-339.

[12] Maly, K., Zubair, M., Chilukamarri, V. and Kothari, P. (2005) GRID Based Federated Digital Library, CF'05, Ischia, Itala, pp. 97-105.

[13] Mazzucco, M., Ananthanarayan, A., Grossman, R. L., Levera, J. and Rao, G. B. (2002) Merging Multiple Data Streams on Common Keys over High Performance Networks, Proceedings of the 2002 ACM/IEEE Conference on Supercomputing, Baltimore, Maryland, pp. 1-12.

[14] Paltoglou, G., Salampasis, M. and Satratzemi, M. (2007) Hybrid results merging, CIKM'07, New York, NY, USA, pp. 321-330.

[15] Paltoglou, G., Salampasis, M. and Satratzemi, M. (2008) Integral Based Source Selection for Uncooperative Distributed Information Retrieval Environments, LSDS-IR'08, Napa Valley, California, pp. 67-74.

[16] Rao, R. (2004) From IR to Search, and Beyond, Queue: Open Source Grows Up, 2, 66-73.

[17] Shokouhi, M., Baillie, M. and Azzopardi, L. (2007) Updating Collection Representations For Federated Search, SIGIR'07, Amsterdam, The Netherlands, pp. 511-518.

[18] Shokouhi, M. and Zobel, J. (2007) Federated Text Retrieval From Uncooperative Overlapped Collections, SIGIR'07, Amsterdam, The Netherlands, pp. 495-502.

[19] Si, L. and Callan, J. P. (2002) Using Sampled Data adn Regression to Merge Search Engine Results, SIGIR'02, Tempere, Finland, pp. 19-26.

[20] Si, L. and Callan, J. P. (2003) A semisupervised learning method to merge search engine results, ACM Transaction of Information Systems, 21, 457-491.

[21] Si, L. and Callan, J. P. (2005) Modeling Search Engine Effectiveness for Federated Search, SIGIR'05, Salvador, Brazil.

[22] Simon, G. and Bird, S. (2003) Building an Open Language Archives Community on the OAI Foundation, Library High Tech, 21, 210-218.

[23] Song, M. and Bieber, M. (2008) IntegraL: Lightweight Link-Based Integration of Heterogeneous Digital Library collections and Services in the Deep Web, 10th IEEE Conference on E-Commerce Technology and the 5th IEEE Conference on Enterprise computing, E-Commerce and E-Services pp. 369-375.

[24] Wu, Y. B., Li, Q., Bot, R. S. and Chen, X. (2006) Finding Nuggets in Documents: A Machine Learning Approach, Journal of the American Society For Information Science and Technology, 57, 740-752.

Music and Mood: Where Theory and Reality Meet

Xiao Hu
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 E. Daniel St. Champaign, IL, 61820
xiaohu@illinois.edu

ABSTRACT
The affective aspect of music, often referred as music mood or emotion, has been recently recognized as an important factor in organizing and accessing music information. However, music mood is far from being well studied in information science. For example, there is no consensus on whether to use *mood* or *emotion* to refer the affective aspect of music. Also, the lack of consensus on music mood categories in the Music Information Retrieval (MIR) community makes it difficult to compare classification approaches developed in different laboratories. On the other hand, there is a rich literature in music psychology that has addressed many of the issues MIR researchers want to know. This research reviews theories in music psychology and summarizes fundamental insights that can help MIR researchers in interpreting music mood. In order to investigate whether classic theories are still applicable to today's reality of music listening environment, this study also derives a set of music mood categories from social tags, using a combination of linguistic resources and human expertise, and compares it to music mood categories in psychological theories. The results verify that there are common grounds between theoretical music mood models and the reality of music listening, but theoretical models do not cover all mood categories emerged from social tags and thus need to be modified to better fit the reality of music listening.

Categories and Subject Descriptors
H.3.7 [**Information Storage and Retrieval**]: Digital Libraries – *standards, user issues.* J.5 [**Computer Applications**]: Arts and Humanities – *music*

General Terms
Human Factors, Standardization, Theory, Verification.

Keywords
Music, mood, metadata, social tags, music psychology, emotion theories, music mood categories, music information retrieval

1. INTRODUCTION
Perhaps no one, be he a music expert or casual listener, would deny the fact that music and mood can never be separated. Some music may not describe a story, but all music must express, strongly or softly, a certain emotion or a mixture of emotions. In consequence, music listeners often experience some sort of affective responses. Just as Juslin and Sloboda [15] stated:

"Some sort of emotional experience is probably the main reason behind most people's engagement with music. Emotional aspects of music should thus be at the very heart of musical science."

Nevertheless, the affective aspects of music have just started drawing attention in information science in recent years when user studies discovered that music mood is an important factor in music information seeking and organization [5][18][33]. In the Music Information Retrieval (MIR) and Music Digital Libraries (MDL) community, there are many fundamental issues on music mood remaining unresolved. For example, there is no terminology consensus on the very topic we are studying: some researchers use "music emotion", some others use "music mood" to refer the affective aspects of music. On the other hand, there is a long history of influential studies in music psychology where these issues have been well studied. Hence, MIR researchers and information scientists who are interested in music mood should learn from music psychology literature on theoretical issues such as terminology and sources of music mood.

However, not all parts of psychological theories can be borrowed into MIR research because most studies in music psychology were conducted in laboratory settings while today's music listening environment has rich social context brought by the flourishing of Web 2.0. For instance, in studying music mood classification techniques, MIR researchers have employed some influential music mood models such as Russell's two-dimensional music emotion model [25] and Watson's two level hierarchical model [34]. There are two problems in adopting various psychological models: 1) although these models have good theoretical roots, they generally lack the social context of music listening [14]. It is unknown whether the models can well fit today's reality; 2) the lack of consensus on music mood categories makes it hard to compare different automatic classification approaches. Therefore, this study strives to identify music mood categories from social tags that reflect the reality of music listening, and compare the categories to those in theoretical models.

The rest of this paper is organized as follows: Section 2 reviews and summarizes important findings in representative studies on music and mood in music psychology. In Section 3, we describe a method of deriving music mood categories from social tags. A detailed comparison between music mood categories in psychological models and those found in social tags is presented in Section 4. We then draw conclusions in Section 5.

2. THE THEORIES
2.1 Mood vs. Emotion
Since the early stage of music psychologiyl studies, researchers have paid attention to clarifying the concepts of *mood* and *emotion*. The most influential first work formally analyzing music and mood using psychological methodologies is probably Meyer's *Emotion and Meaning in Music* [23]. In this book, Meyer stated that *emotion* is "temporary and evanescent" while *mood* is

"relatively permanent and stable". Sloboda and Juslin [28] followed Meyer's point after summarizing related studies during nearly a half century.

In music psychology, both *emotion* and *mood* have been used to refer to the affective effects of music, but *emotion* seems to be more popular [4][13][23][26][28]. However, in MIR, researchers tend to choose *mood* over *emotion* [7][20][21][24]. In addition, existing music repositories also use *mood* rather than *emotion* as a metadata type for organizing music (e.g., AllMusicGuide[1] and APM[2]). While we have yet to formally interview MIR researchers on why they chose to use *mood*, we hypothesize that there are at least two reasons for MIR researchers to make a different choice from their colleagues in music psychology:

First, as stated by Meyer, *mood* refers to a relatively long lasting and stable emotional state. While psychologists emphasize on human responses to various stimuli of emotion, MIR researchers, at least at current stage, are more interested in the general sentiment that music can convey. In another word, music psychologists focus on the very subjective responses to music which can be acute, momentary and fast changing, while the MIR community tries to find the common affective consequences of music that are shared by many people and are less volatile.

Second, the research purposes of the two disciplines are different. Music psychologists want to discover why a human has emotional responses to music while MIR researchers want to find a new metadata type to organize and access music objects. The former focuses on human's responses, the latter focuses on music. It is human who has *emotion*. Music does not have emotion, but it can carry a certain *mood*.

Therefore, this research continues the choice of MIR researchers and adopts the term *music mood* rather than *emotion*. However, it is noteworthy that the two concepts are not absolutely detached. To some extent, their difference mainly lies in granularity. MIR researchers can still borrow insights from music psychology studies. In fact, when MIR technologies are developed to a level where individual and transitory affective responses become the subject of study, it is possible that the MIR community may change to adopt the notion of music *emotion*.

2.2 Sources of Music Mood

Where music mood comes from is a question MIR researchers are interested in. Does it come from the intrinsic characteristics of music pieces or from the extrinsic context of music listening behaviors? The answer to this question would have significant implications on assigning mood labels to music pieces either by hand or by computer programs.

From as early as Meyer [23], there have been two contrasting views of music meanings in music psychology: the absolutist versus referentialist views. The absolutist view claimed "musical meaning lies exclusively within the context of the work itself" while the referentialist proposed "musical meanings refer to the extra-musical world of concepts, actions, emotional states, and character". Meyer acknowledged the existence of both types of musical meanings. Later, Sloboda and Juslin [28] echoed Meyer's view by presenting two sources of emotion in music: intrinsic emotion and extrinsic emotion. Intrinsic emotion is triggered by specific structural characteristics of the music while extrinsic emotion is from the semantic context related but outside the music. Therefore, the suggestion for MIR is that music mood should be a combination of music content itself and the social context where people listen to and share opinions about music. In fact, recent user studies in MIR have confirmed this point of view (e.g., [18]) and automatic music categorization systems (e.g., [2]) have started to combine music content (e.g., audio, lyrics, and symbols) and context (e.g., social tags, playlists, and reviews).

2.3 What We Know about Music Mood

Beside terminology and sources of music mood, music psychology studies on music mood have a number of fundamental generalizations that can benefit MIR research.

1. There does exist mood effect in music. Ever since early experiments (pre-1950) on psychological effects of music, studies have confirmed the existence of the functions of music in changing people's mood [4]. It is also agreed that it seems natural for listeners to attach mood labels to music pieces [28].

2. Not all moods are equally likely to be aroused by listening to music. In a study conducted by Schoen and Gatewood [26], human subjects were asked to choose from a pre-selected list of mood terms to describe their feelings while listening to 589 music pieces. Among the presented moods, sadness, joy, rest, love, and longing were among the most frequently reported while disgust and irritation were the least frequent ones.

3. There do exist uniform mood effects among different people. Sloboda and Juslin [28] summarized that listeners are often consistent in their judgment about the emotional expression of music. Early experiments in [26] have shown that "*the moods induced by each (music) selection, or the same class of selection, as reported by the large majority of our hearers, are strikingly similar in type*". Such consistency is an important ground for developing and evaluating music mood classification techniques.

4. Not all types of moods have the same level of agreement among listeners. Schoen and Gatewood [26] ranked joy, amusement, sadness, stirring, rest and love as the most consistent moods while disgust, irritation and dignity were of the lowest consistency. The implication for MIR is that some mood categories would be harder to classify than others.

5. There is some correspondence between listeners' judgments on mood and musical parameters such as tempo, dynamics, rhythm, timbre, articulation, pitch, mode, tone attacks and harmony [28]. Early experiments showed that the most important music element for excitement was swift tempo; modality was important for sadness and happiness but useless for excitement and calm; and melody played a very small part in producing a given affective state [4]. Schoen and Gatewood [26] pointed out the mood of amusement largely depended upon vocal music: "*humorous description, ridiculous words, peculiarities of voice and manner are the most striking means of amusing people through music*". This has been evidenced by the category, "*humorous/silly/quirk*" used in the Audio Mood Classification (AMC) task in the Music

[1] http://allmusic.com

[2] http://www.apmmusic.com

Information Retrieval Evaluation eXchange (MIREX)[3], a formal evaluation framework in the MIR community [12]. A subsequent examination on the AMC data found that music pieces which were manually labeled with this category mostly had the above mentioned quality. Such correspondence between music mood and musical parameters has very important implications for designing and developing music mood classification algorithms.

2.4 Music Mood Categories

Studies in psychology have proposed a number of models on human's emotions and music psychologists have adopted and extended a few influential models.

The six "universal" emotions defined by Ekman [6]: anger, disgust, fear, happiness, sadness, and surprise, are well known in psychology. However, since they were designed for encoding facial expressions, some of them may not be suitable for music (e.g., disgust), and some common music moods are missing (e.g., calm or soothing). In music psychology, the earliest and still best-known systematic attempt at creating music mood taxonomy was by Hevner [10]. Hevner designed an adjective circle of eight clusters of adjectives as shown in Figure 1, from which we can see: 1) the adjectives within each cluster are close in meaning; 2) the meanings of adjacent clusters would differ slightly; and 3) the difference between clusters gets larger step by step until a cluster at the opposite position is reached.

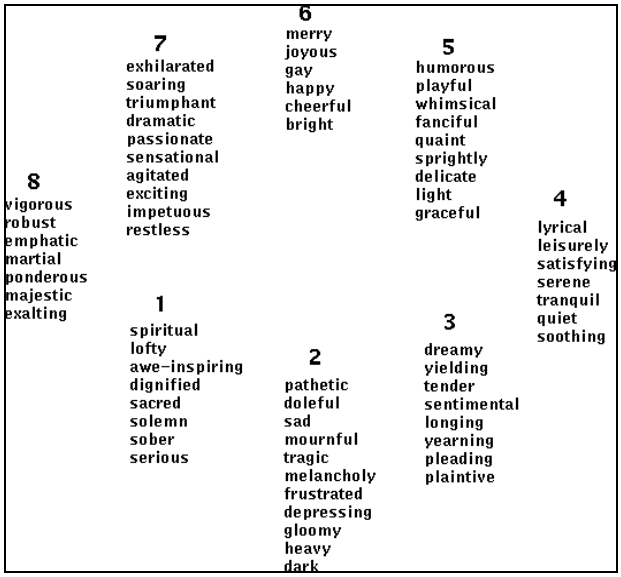


Figure 1: Hevner's adjective cycle [10].

Both Ekman's and Hevner's models belong to *categorical* models because the mood spaces consist of a set of discrete mood categories. Another well recognized kind of models is *dimensional* models where emotions are positioned in a continuous multidimensional space. The most influential ones contain such dimensions as Valence (happy-unhappy), Arousal (active-inactive), and Dominance (dominant-submissive) [22][25][31]. However, there is no consensus on how many dimensions there should be and which dimensions to consider. For example, a well cited study by Wedin identified three dimensions: Intensity-Softness, Pleasantness-Unpleasantness and Solemnity-Triviality [35], while another study by Asmus found nine dimensions: Evil, Sensual, Potency, Humor, Pastoral, Longing, Depression, Sedative, and Activity [1].

Among all these dimensional models, the Russell's model of the combination of valence and arousal dimensions [25][31] has been adopted in a few experimental studies in music psychology (e.g., [27][32]), and MIR researchers have been using similar taxonomies based on this model (e.g., [16][17][20]). As shown in Figure 2, the original Russell's model places 28 emotion denoting adjectives on a circle in a bipolar space consisting of valence and arousal dimensions.

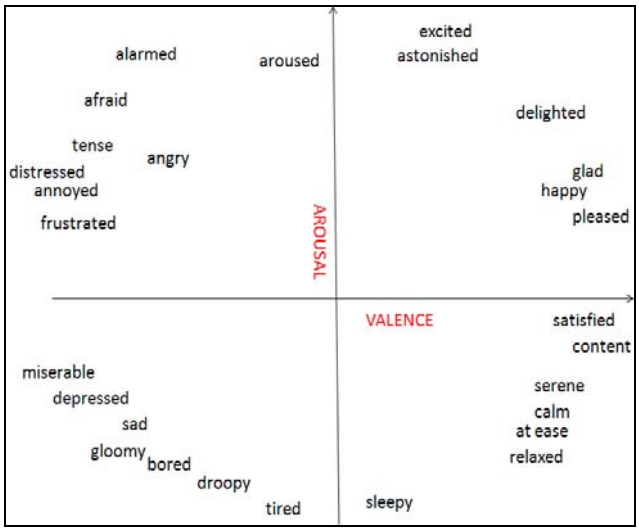


Figure 2: Russell's model with two dimensions: arousal and valence [25].

In fact, categorical models and dimensional models cannot be completely separated. Gabrielsson and Lindström [8] argued that Hevner's model suggested an implicit dimensionality similar to the combination of valence (cluster 2 – cluster 6) and arousal (clusters 7/8 – clusters 4/3).

All these psychological models were proposed in laboratory settings and thus were criticized as being lack of social context of music listening [14]. In the next section, we will derive a set of music mood categories from social tags, and in Section 4 we will compare it to those in the Hevner's model as well as the Russell's model.

3. THE REALITY

3.1 Mood Categories in Social Tags

With the birth of Web 2.0, the general public can now post text tags on music pieces and the large quantity of social tags become a unique and rich resource of discovering users' perspectives in the social context of music listening. MIR Studies have tried to find music mood or genre representations from social tags (e.g., [11][19]), but none of them have adequately addressed the following shortcomings of social tags as summarized by Guy and Tonkin [9]. First, social tags are uncontrolled and thus contain much noise or junk tags. Second, many tags have ambiguous meanings. For example, "love" can be the theme of a song or a user's attitude towards a song. Third, a majority of tags are tagged

[3] http://www.music-ir.org/mirex/2007/index.php/AMC

to only a few songs, and thus are not representative (i.e., the so called "long-tail" problem). Fourth, some tags are essentially synonyms (e.g., "cheerful" and "joyful"), and thus do not represent separate and distinguishable categories. To address these problems, we propose a new method that combines the strength of linguistic resources and human expertise to derive more realistic and user-centric mood categories from social tags.

3.1.1 Identifying mood-related terms

First, we identified a set of mood related terms using linguistic resources. WordNet-Affect is an affective extension of WordNet [30]. It assigns affective labels to words representing emotions, moods, situations eliciting emotions, or emotional responses. As a major resource used in text sentiment analysis, WordNet-Affect has a good coverage of mood related words. There are 1,586 unique terms in WordNet-Affect. However, some of the terms are judgmental, such as "bad", "poor", "miserable", "good", "great", and "amazing". Although these terms are related to mood, their applications on songs probably represent users' judgments towards the songs, rather than describe the moods carried by the songs. Therefore, such tags are noise for our purposes and should be eliminated. Another linguistic resource, General Inquirer [29] was consulted for a list of judgmental terms. General Inquirer is a lexicon comprised of 11,788 words organized in 182 psychological categories, two of which are about "evaluation" containing 492 words implying judgment and evaluation. Subtracting these words from terms in WordNet-Affect resulted in1,384 terms.

As a final step to ensure the quality of the term list, two human experts were consulted and manually examined the terms. Both experts are MIR researchers with a music background and native English speakers. They first identified and removed tags with music meanings that did not involve an affective aspect (e.g., "trance" and "beat"). Then, they removed words with ambiguous meanings. For example, "chill" can mean "to calm down" or "depressing", but social tags do not provide enough contexts to disambiguate the term. After this step, we got 1,249 mood related terms.

3.1.2 Obtaining mood-related social tags

Last.fm is one of the most popular tagging sites for Western music[4]. With 30 million users every month, it provides a good resource of studying how people tag music. We queried last.fm through its API[5] with the 1,249 mood related terms, and 476 of them have been used as tags by last.fm users as of June 2009. To untangle the "long-tail" problem mentioned above, we only included tags that were used more than 100 times. This gave us 146 terms/tags.

3.1.3 Grouping mood-related social tags

To solve the synonym problem of social tags, we grouped the 146 mood related tags such that synonyms were merged together into one category. We again used WordNet-Affect in this step. WordNet is a natural resource for identifying synonyms, because it organizes words into *synsets*. Words in the same synset are synonyms in the linguistic point of view. Moreover, WordNet-Affect also links each non-noun synset (verb, adjective and adverb) with the noun synset from which it is derived. For instance, the synset of "joyful" is marked as derived from the synset of "joy". Both synsets represent the same kind of mood and should be merged into the same category. Hence, mood-related tags appearing in and being derived from the same synset in WordNet-Affect were merged into one group.

Finally, human experts were again consulted to modify the grouping of tags when they saw the need of splitting or further merging some groups. As a result, 36 categories emerged. Table 1presents some of them major categories and the number of tags contained in each category[6].

Table 1: Major mood categories derived from last.fm tags

Categories	#. tags
calm, calm down, calming, calmness, comfort, quiet,..	16
gloomy, blue, dark, depress, depressed, depressing,…	10
mournful, grief, heartache, heartbreak, heartbreaking,…	9
cheerful, cheer up, cheer, cheery, festive, jolly, merry,…	8
gleeful, euphoria, euphoric, high spirits, joy, joyful,…	8
brooding, broody, contemplative, meditative, pensive,…	7
confident, encouragement, encouraging, fearless,…	6
exciting, exhilarating, stimulating, thrill, thrilling	5
anxious, angst, anxiety, jumpy, nervous	5
angry, anger, furious, fury, rage	5
compassionate, mercy, pathos, sympathy	4
desolate, desolation, isolation, loneliness	4
scary, fear, panic, terror	4
hostile, hatred, malevolent, venom	4
glad, happiness, happy	3
hopeful, desire, hope	3
sad, melancholic, sadness	3
aggression, aggressive	2
romantic	1
surprising	1

4. COMPARISONS ON MOOD CATEGORIES

The social tagging environment of Web 2.0 is very different from the laboratory settings where the music psychology studies were conducted. Hence it is interesting to compare the mood categories derived from social tags to the models developed in music psychology. Such comparison will disclose whether the theoretical models can support patterns emerged from empirical data and how much differences are between them. Specifically, the following questions are addressed:

(1) Is there any correspondence between the resultant categories and those in the psychological models?

(2) Do the distances between mood categories show similar patterns to those in the psychological models?

Both Hevner's categorical model and Russel's two-dimensional model are compared to the derived categories.

[4] http://socialmediastatistics.wikidot.com/lastfm Retrieved at July 22, 2008.

[5] http://www.last.fm/api

[6] Due to space limit, the complete list can be found at http://www.isrl.illinois.edu/~xiaohu/pub/iconf10/Table1.pdf

4.1 Categories

4.1.1 Hevner's circle vs. derived categories

Some of the terms in Hevner's circle (Figure 1) are known to be old-fashioned and are rarely used for describing moods nowadays. This is reflected by the fact that only 37 of the 66 words in Hevner's circle were found in WordNet-Affect, including matches of terms in different derived forms (e.g., "solemnity" and "solemn" were counted as a match). Comparing the clusters in Hevner's circle to the set of categories identified from social tags, we found that 33 words (50% of all) in Hevner's circle matched tags in the derived categories, as indicated in Figure 3 where matched words are surrounded by rectangles. Please note that in Figure 3 the order of words within each cluster may be changed from Figure 2, so that words in the same derived categories are within one rectangle. The observation that the rectangles never cross Hevner's clusters suggests that the boundaries of Hevner's clusters and derived categories are in accordance to each other, despite the derived categories are of a finer granularity.

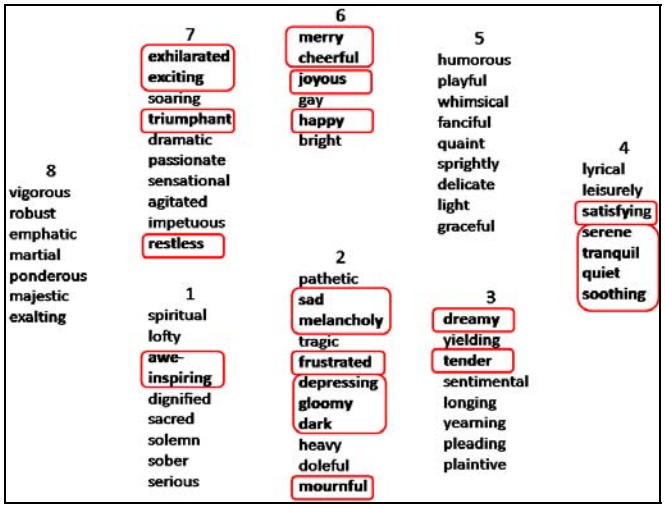


Figure 3: Words in Hevner's circle that match tags in the categories derived from last.fm tags

From Figure 3, we can also see that Clusters 2, 4, 6, 7 have the most matched words among all clusters, indicating Western popular songs (as the main music type in last.fm) mostly fall into these mood clusters. Besides exact matches, there are five categories in Table 1 with meanings close to some of the clusters in Hevner's model: categories "angry", "aggressive" are close to Cluster 8, category "desire" is close to "longing" and "yearning" in Cluster 3, and category "earnest" is close to "serious" in Cluster 1. This use of different words for the same or similar meanings indicates a vocabulary mismatch between social tags and adjectives the Hevner's model. Clusters 1 and 5 have the least matched or nearly matched words, reflecting that they are not good descriptors for Western popular songs. In fact, Hevner's circle was mainly developed for classical music for which words in Clusters 1 and 5 ("light", "dedicate" and "graceful") would be a good fit.

In total, 20 of the 36 derived categories have at least one tag contained in Hevner's circle. This is not surprising that empirical data entailed more categories since social tags were aggregated from millions of users while Hevner's model was developed by studying hundreds of subjects.

As a conclusion, after more than seven decades, Hevner's circle is still largely in accordance to categories derived from today's empirical music listening data. Admittedly, there are more mood categories in today's reality and there is a vocabulary mismatching issue, since language itself is evolving with time.

4.1.2 Russell's model vs. derived categories

Figure 4 marks the words appearing in both Russell's model and the derived sets of mood categories. Words in the same category are circled together.

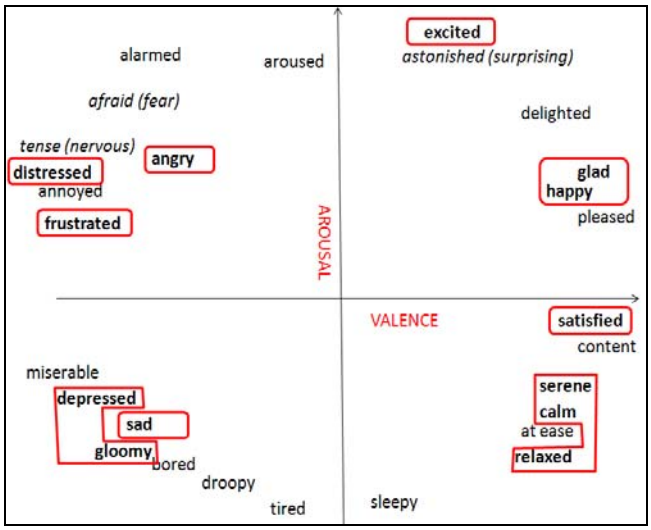


Figure 4: Words in Russell's model that match tags in the categories derived from last.fm tags

Figure 4 shows that 13 of the 28 words in Russell's model match tags in the derived categories (marked in bold), and another 3 words (marked in italic) have close meanings with tags in the derived categories (shown in parentheses). Hence, more than half of the words in Russell's model match or nearly match tags in the derived categories. For those unmatched words, there are several cases: 1) Some words in the Russell's model are synonyms according to WordNet, such as "content" and "satisfied"; "at ease" and "relaxed"; "droopy" and "tired", "pleased" and "delighted". Words in these pairs represent similar mood. 2) Some words are ambiguous and can be judgmental ("miserable", "bored", "annoyed"). If used as social tags, these terms may represent users' preferences towards the songs rather than the moods carried by the songs. Hence these terms were removed during the process of deriving mood categories from social tags. 3) 5 of the 28 adjectives in Russell's model are not in WordNet-Affect: "aroused", "tense", "droopy" and "sleepy". They are either rarely used in daily life or are not deemed as mood-related. Nevertheless, the high percentage of matched vocabulary with WordNet-Affect (23 out of 28) does reflect the fact that Russell's model is newer than Hevner's.

We can also see from Figure 4 that matched words in the same category (circled together) are placed closely in the Russell's model, and the matched words distribute evenly across the four quadrants of the two dimensional space. This indicates the derived categories have a good coverage of moods in the Russell's model. On the other hand, 2/3 of the 36 derived categories do not have matched words in the Russell's model. Therefore, this comparison tells us that the Russell model

simplifies the problem in reality and the MIR experiments based on this model did help classify *some* of the mood categories used in real life but not *all* of them. Nevertheless, let us recall that Russell's model is a dimensional model instead of a categorical model, and thus theoretically it is not limited to the 28 adjectives. In fact, later studies have extended this model in many different ways [27][31][32]. It is possible (with further verifications in psycholinguistics) that most, if not all tags in the derived categories could find their places in the two-dimensional space, but it is a topic of beyond the scope of this paper.

4.2 Distances between Categories

Both Hevner's circle and Russell's space demonstrate relative distances between moods. For instance, in Russell's space, "sad" and "happy", "calm" and "angry" are at opposite places while "happy" and "glad" are close to each other.

To see if there are similar patterns in the derived categories, we calculated the distances between them. Last.fm API provides top 50 artists associated with each tag. We collected top artists for each of the 146 tags in the derived categories and calculated distances between the categories based on artist co-occurrences. Figure 5 shows the distances of the sets of categories plotted in a 2-dimensional space using Multidimensional Scaling [3].

As shown in Figure 5, categories that are intuitively close (e.g., those denoted by "glad", "cheerful", "gleeful") are positioned together, while those placed at almost opposite positions indeed represent contrasting moods (e.g., the ones denoted as "aggressive" and "calm", "cheerful" and "sad"). This evidences that the mood categories derived from social tags have similar patterns of category distances to those in psychological mood models.

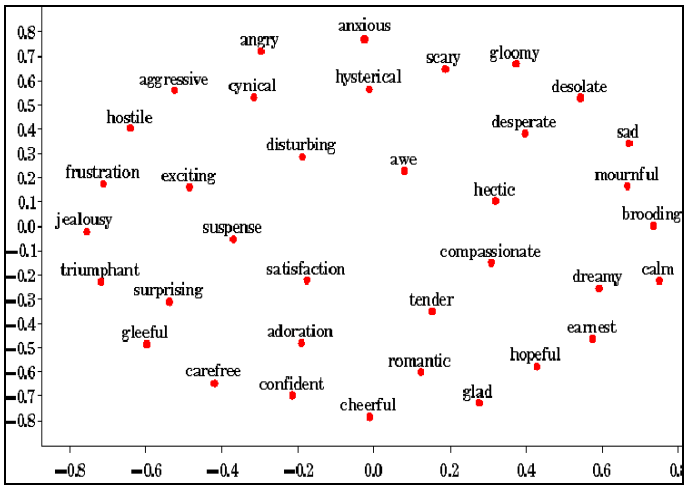


Figure 5: Distances of the 36 derived mood categories based on artist co-occurrences (each category is denoted by one tag in that category)

From the above comparisons, we can see that the derived set of categories is in accordance to common sense and is at least partially supported by classic psychological models. In addition, the derived categories are more comprehensive than psychological models and are more closely connected with the reality of music listening. It is our suggestion and recommendation for the MIR/MDL community to adopt the derived category set in music mood classification experiments, which will also facilitate comparisons across approaches.

5. CONCLUSIONS

Music mood is a newly emerged metadata type of music. Researchers in the MIR/MDL community have a lot to learn from psychology literature, from basic terminology to music mood categories. This paper reviews seminal works in the long history of psychological studies on music and mood, and summarizes fundamental points of view and their important implications on MIR/MDL research.

In MIR, one of the most debated topics on music and mood is mood categories. Theoretical models in psychology were designed from laboratory settings and may not be suitable for today's reality of music listening. By deriving a set of mood categories from social tags and comparing it to the two most representative mood models in psychology, this study finds out there are common grounds between theoretical models and categories derived from empirical music listening data in the real life. On the other hand, there are also non-neglectable differences between categories in theory and those in reality: 1) Vocabularies are different. Some words used in theoretical models are outdated, or otherwise not used in today's daily life; 2) Targeted music is different. Theoretical models were mostly designed for classical music while there are a variety of music genres in today's music listening environment; 3) While theoretical models often have a handful number of mood categories, the reality can have more categories and in a finer granularity. Therefore, in developing music mood classification techniques for today's music and users, MIR researchers should extend classical mood models according to the context of targeted users and music listening reality. For example, to classify Western popular songs, Hevner's circle can be adapted by introducing more categories found from social tags and trimming Cluster 1 and 5 which are mostly for classical music.

Information science is an interdisciplinary field. It often involves topics that have been traditionally studied in other fields. Borrowing findings from literatures in other fields is a very important research method in information science, but we need to pay attention to connecting theories in the literature to the reality and social context of the problems we investigate. The study described in this paper is a good example of connecting music psychology literature to the reality of music listening in the context of studying music mood as a new metadata type of music. In general, the methodology of literature review, analysis on empirical data and comparison of the two can help information science researchers refine or adapt theoretical models to better fit the reality of users' information behaviors.

6. ACKNOWLEDGEMENT

We thank the Andrew W. Mellon Foundation for their financial support. We also thank the anonymous reviewers for their helpful comments and suggestions.

7. REFERENCES

[1] Asmus, E. 1995. The development of a multidimensional instrument for the measurement of affective responses to music. Psychology of Music, 13(1): 19-30.

[2] Aucouturier, J-J., Pachet, F., Roy, P. and Beurivé, A. 2007. Signal + Context = Better Classification. In Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07). Sept. 2007, Vienna, Austria.

[3] Borg, I. and Groenen, P. J. F. 2004. Modern Multidimensional Scaling: Theory and Applications. Springer.

[4] Capurso, A., Fisichelli, V. R., Gilman, L., Gutheil, E. A., Wright, J. T. and Paperte, F. 1952. Music and Your Emotions. Liveright Publishing Corporation.

[5] Cunningham, S. J., Jones, M., and Jones, S. 2004. Organizing digital music for use: an examination of personal music collections. In Proceedings of ISMIR'04, Barcelona, Spain.

[6] Ekman, P. 1982. Emotion in the Human Face. Cambridge University Press, Second ed.

[7] Feng, Y., Zhuang, Y. and Pan, Y. 2003. Popular music retrieval by detecting mood. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. July 2003, Toronto, Canada.

[8] Gabrielsson, A. and Lindström, E. 2001. The influence of musical structure on emotional expression. In P. N. Juslin and J. A. Sloboda (Eds.), Music and Emotion: Theory and Research. New York: Oxford University Press.

[9] Guy, M. and Tonkin, E. 2006. Tidying up tags. D-Lib Magazine. 12(1). Retrieved from http://www.dlib.org/dlib/january06/guy/01guy.html on November 18, 2008

[10] Hevner, K. 1936. Experimental studies of the elements of expression in music. American Journal of Psychology, 48: 246-68.

[11] Hu, X., Bay, M. and Downie, J. S. 2007. Creating a simplified music mood classification groundtruth set, In Proceedings of ISMIR'07. Sept. 2007, Vienna, Austria.

[12] Hu, X., Downie, J. S., Laurier, C., Bay, M. and Ehmann, A. 2008. The 2007 MIREX Audio Music Classification task: lessons learned. In ISMIR'08. Sept. 2008, Philadelphia, Pennsylvania.

[13] Juslin, P. N., Karlsson, J., Lindström E., Friberg, A. and Schoonderwaldt, E. 2006. Play it again with feeling: computer feedback in musical communication of emotions. Journal of Experimental Psychology: Applied, 12(1): 79-95.

[14] Juslin, P. N. and Laukka, P. 2004. Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. Journal of New Music Research, 33(3): 217-238.

[15] Juslin, P. N. and Sloboda, J. A. 2001. Music and emotion: introduction. In P. N. Juslin and J. A. Sloboda (Eds.), Music and Emotion: Theory and Research. New York: Oxford University Press.

[16] Kim, Y., Schmidt, E., and Emelle, L. 2008. Moodswings: A collaborative game for music mood label collection. In Proceedings of ISMIR'08. Philadelphia, USA.

[17] Laurier, C., Grivolla, J. and Herrera, P. 2008. Multimodal music mood classification using audio and lyrics. In Proceedings of the 7th International Conference on Machine Learning and Applications (ICMLA' 08). December 2008, San Diego, California.

[18] Lee, J. H. and Downie, J. S. 2004. Survey of music information needs, uses, and seeking behaviours: preliminary findings. In Proceedings of ISMIR'04. Barcelona, Spain.

[19] Levy, M. and Sandler, M. 2007. A Semantic Space for MusicDerived from Social Tags, In Proceedings of ISMIR'07, Vienna, Austria.

[20] Lu, L., Liu, D. and Zhang, H. 2006. Automatic mood detection and tracking of music audio signals. IEEE Transactions on Audio, Speech, and Language Processing, 14(1): 5-18.

[21] Mandel, M., Poliner, G. and Ellis, D. 2006. Support vector machine active learning for music retrieval. Multimedia Systems, 12 (1): 3-13.

[22] Mehrabian, A. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology: Developmental, Learning, Personality, Social, 14, 261-292.

[23] Meyer, L. B. 1956. Emotion and Meaning in Music. Chicago: University of Chicago Press.

[24] Pohle, T., Pampalk, E. and Widmer, G. 2005. Evaluation of frequently used audio features for classification of music into perceptual categories. In Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing. June, 2005, Riga, Latvia.

[25] Russell, J. A. 1980. A circumplex model of affect. Journal of Personality and Social Psychology, 39: 1161-1178.

[26] Schoen, M. and Gatewood, E. L. 1927. The mood effects of music. In M. Schoen (Ed.), The Effects of Music (International Library of Psychology) Routledge, 1999.

[27] Schubert, E. 1996. Continuous response to music using a two dimensional emotion space. In Proceedings of the 4th International Conference of Music Perception and Cognition. pp. 263-268.

[28] Sloboda, J. A. and Juslin, P. N. 2001. Psychological perspectives on music and emotion. In P. N. Juslin and J. A. Sloboda (Eds.), Music and Emotion: Theory and Research. New York: Oxford University Press.

[29] Stone, P. J. 1966. General Inquirer: a Computer Approach to Content Analysis. Cambridge: M.I.T. Press.

[30] Strapparava, C. and Valitutti, A. 2004. WordNet-Affect: an affective extension of WordNet. In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)

[31] Thayer, R. E. 1989. The Biopsychology of Mood and Arousal. New York: Oxford University Press.

[32] Tyler, P. 1996. Developing A Two-Dimensional Continuous Response Space for Emotions Perceived in Music. Doctoral dissertation. Florida State University.

[33] Vignoli, F. 2004. Digital Music Interaction concepts: a user study. In Proceedings of ISMIR'04. Barcelona, Spain.

[34] Watson, D. 2000. Mood and Temperament, Guilford Press, New York, NY, USA,

[35] Wedin, L. 1972. A multidimensional study of perceptual-emotional qualities in music. Scandinavian Journal of Psychology, 13: 241–257.

Exploring Collaborative Rhythm: Temporal Flow and Alignment in Collaborative Scientific Work

Steven J. Jackson
University of Michigan
1075 Beal Ave, Rm 3212
Ann Arbor, MI 48109
(011)-734-764-8058
sjackso@umich.edu

David Ribes
Georgetown University
3520 Prospect St NW, Suite 311
Washington, DC 20057
(011)-202-687-4831
dr273@georgetown.edu

Ayse Buyuktur
University of Michigan
1075 Beal Ave
Ann Arbor, MI 48109
(011)-734-
abuyuktu@umich.edu

ABSTRACT
Recent studies of large-scale distributed practice in the sciences and elsewhere have taught us important things about space and place as props and barriers to distributed collective action, but they have had relatively less to say about time. This paper offers a typology of collaborative rhythms and argues for the importance of *temporal alignment* as a neglected but crucial element underpinning distributed collective practice in the sciences (and we believe other spheres of distributed collective activity). Specifically, we argue that joint scientific work is organized around four separate and potentially dissonant temporal registers, or 'rhythms' – *phenomenal*, *organizational*, *biographical*, and *infrastructural* – and that efforts to align such rhythms constitute an important and under-recognized aspect of collaborative work. The ideas and examples are drawn from the authors' own field studies around IT infrastructure ('cyberinfrastructure') and collaborative practices across a range of scientific fields.

General Terms
Management, Design, Human Factors, Standardization, Theory.

Keywords
Time, rhythm, collaboration, science, cyberinfrastructure, ethnography.

1. INTRODUCTION
Recent studies of large-scale collaboration in the sciences have taught us important things about space and place as props and barriers to distributed collective action, but they have had relatively little to say about time. This paper offers a typology of collaborative rhythms and the ongoing work of temporal alignment as neglected but crucial elements underpinning distributed collective practice in the sciences (and we believe, many other spheres of distributed collective activity). The ideas and examples are drawn from our own field studies – joint and individual, past and current – around IT infrastructure ('cyberinfrasructure') and collaborative practices across a range of scientific fields (Ribes 2006; Edwards, Jackson et al. 2007; Jackson, Edwards et al. 2007; Ribes and Finholt 2007; Ribes and Bowker 2008).

As Lakoff and Johnson have argued (Lakoff and Johson 1980), far from a figurative add-on to the basic business of cognition, metaphor structures our fundamental categories of thought and expression. Time, we are told, is a river, by which is usually meant that it flows uniformly and ineluctably forward. But this metaphor may be richer than we know, for as any white water canoeist or first year hydrology student will tell you, one of the more fascinating, and theoretically challenging characteristics of rivers is that they flow at many different speeds – and even in many different directions – at once. More importantly, part of managing collaborative rhythms is the work of temporal alignment, bringing heterogeneous patterns in synch for moments of coordinated activity. Following Heraclitus, it true that you can never step in the same river twice, but the complex eddies and whorls of streams, in combination with our human dams and levies, paints an image of time more complex both in its nature and in our ability to act upon it. In this paper, we will argue that the rhythm or timeliness of collaborative scientific work has this blended, layered, and every-which-way-at-once quality while also being the active object of our efforts to bring it under control– and that this fact has been routinely neglected in the study of large-scale scientific collaborative work to date.

2. TIME AND COLLABORATION
Research on distributed collective work in recent years has paid considerable attention to the variable effects of distance and spatial location on collaborative form and practice, including their effects on collaborative outcomes – i.e., 'success' and 'failure' as measured along definable parameters like publication rates, co-authorship patterns, and other markers of collaborative productivity (Cummings and Kiesler 2007). This research has led information and other scholars towards more nuanced and specific understandings of the ways in which spatial constraints and affordances may shape and condition the nature of distributed work. Much of this work has focused on the secret assists that shared place provides in the structuring of collaborative activity: from its effects on contextual awareness and interpersonal trust (Kiesler and Cummings 2002; Schmidt 2002) to its role in sustaining group-level identities, mediating conflict, and building effective common ground (Clark and Brennan 1991; Hinds and Mortenson 2005). In response, much of the work coming out of the design wing of the CSCW and HCI communities has been about recreating the hidden affordances of place in now distributed technical and organizational forms, seeking to restore through design the ever-elusive experience of "being there"(Hollan and Stornetta 1992). In broad keeping with the 'spatial turn' in the social sciences at large, we now generally acknowledge that in the design and practice of large-scale collaborative organizations, "distance matters"(Olson and Olson 2000).

The same cannot be said about our thinking around time in such settings, the study of which remains rudimentary, fragmented, and

both theoretically and empirically under-analyzed. A recent review of key journals in the organizational science and computer-supported cooperative work (CSCW) fields reveals a relative dearth of articles tackling questions of temporality and collective work in serious and sustained ways. There is a literature that focuses on incongruent temporal rhythms that arise from collaborators working in different time zones, usually in inter-continental work teams of transnational corporations, where temporal differences are reduced to side effects of distance. Other works on time and distributed teams distinguish between synchronous and asynchronous communication among team members, discussing the ways in which these support collaborative activities. Aspects of synchronicity are also often discussed in the context of affordances of different communication and information technologies that support collaborative work. However, most studies have treated temporal issues in rather narrow scope, focusing on one facet without paying attention to the many different and fluctuating rhythms present in collaborative work. There is a relative scarcity, for example, of efforts to incorporate social rhythms into discussions of distance collaboration.

This general absence sits against the backdrop of a renewed and growing interest in social theory and the social sciences at large. Beginning in the 1980s (but reviving themes as old as Marx), scholars became interested again in the reciprocal effects of large-scale social and economic restructuring and the distinctive experiences of speed (Virilio 1986) and time-space compression (Harvey 1991) or 'distanciation' (Giddens 1991) that marked and structured the social forms of "modernity" (late, post, second, etc.). More recent work has sought to extend and nuance this analysis, introducing various and revivified forms of time geography, some building from traditions of geographic research dating to the 1950s. Heroes of the spatial turn such as Henri Lefebvre have returned in later life to consider the under-articulated temporalities implicit in the spatial forms of things, likening the Mediterranean city to a sort of time machine built around the compression and coordination of historical rhythms of variable kinds and periodicities; such studies constituted fragments and beginning points for a larger project, unfinished at the time of his death, that Lefebvre referred to (but never fully described) as "rhythmanalysis" (Lefebvre 2004). Other sources for the revival of temporal thinking in social theory have been drawn from the field of history, most notably the work of the French 'Annaliste' historians and their efforts to mark both distinctions and connections between histories of the short, medium, and long 'durees'(Braudel 1992; Braudel 2004). Still others have been inspired by linguistics and literary theory, including Foucauldian and Bakhtinian-inspired ideas around 'pluritemporalism' or biologically inspired examples around 'heterochronicity' (Nowotny 1992; Lemke 2000). Together these explore the coexistence of multiple modes or registers of time in the structure and practice of ongoing social activity and point to the difficulty of coordinated time across institutions, professional bodies and career trajectories; unfortunately they each ignore the phenomenal rhythms so key to the Annalists. In this way they represent an unfortunate branch of studies of new forms of science which entirely black-box the domain of science (and its objects of study).

In organizational science, early work by Barley (Barley 1988) and more recent work by Orlikowski and Yates (Orlikowsi and Yates 2002) has made forceful arguments around the 'enacted' character of time and its relationship with organizational form and practice. Orlikowksi and Yates take particular issue with the long-standing theoretical split between objective ('clock time') and subjective ('event time') understandings of temporality in organizational practice. They note that

> difficulties arise when these positions are treated – not as conceptual tools – but as inherent properties of time. Focusing on one side or the other misses seeing how temporal structures emerge from and are embedded in the varied and ongoing social practices of people in different communities and historical periods, and at the same time how such temporal structures powerfully shape those practices in turn. (686)

From this classically structurational perspective, time appears as both medium and outcome of ongoing social practice, simultaneously shaping and shaped by the choices of human actors. One important advantage of this perspective comes with the seriousness it accords individual and group-level choices in altering the temporal forces that would otherwise appear to impinge on them very much from the outside; from this perspective

> people are purposive, knowledgeable, adaptive, and inventive actors who, while they are shaped by established temporal structures, can also choose (whether explicitly or implicitly) to (re)shape those temporal structures to accomplish their situated and dynamic ends. (688)

The same principle supplies an account of temporal change, and reminds us of the potentially fragile nature of apparently objective or 'timeless' temporal orders. For Orlikowski and Yates, temporal structuring also provides a vehicle for talking across a series of entrenched divides – universal vs. particular, linear vs. cyclical, natural vs. social, open-ended vs. closed – that have hobbled social scientific research on time to date.

Broadly parallel interests can be found in recent CSCW work by Bardram (Bardram 2000) and Reddy, Dourish, and Pratt (Reddy, Dourish et al. 2006). Like Orlikowski and Yates (and in broad sympathy with their critique of objectivism), these authors explore "the production and negotiation of temporal order… as a practical accomplishment of social actors" (Reddy et. al. 31). In particular, they seek to account for the temporal organization of work in the surgical settings they study as the outcome of three central features: temporal *trajectories* (focused on the illness trajectories of individual patients); temporal *rhythms* (manifested in repeated patterns of work at the collective level); and temporal *horizons* (roughly, the ways in which individuals order and orient their work within the constraints of broader organizational rhythms). Having articulated such features of organizational time, the authors conclude with an argument for building time-sensitive notions of flow and awareness into the conceptualization and design of medical information spaces.

These early forays of organizational science and CSCW into time haven't been taken up in a robust way by the community at large. Our immediate concern is that the centrality of time and rhythmic alignment to collaborative practices of all sorts has yet to be charted in a place we see these effects turning up in interesting and surprising ways: the practice of large-scale collaborative science. Building from our own studies of distributed collective practice across a range of scientific fields, this paper explores the

inherent and diversely-constituted timeliness of collaborative work and the practical barriers this diversity may pose, as well as pointing to the distinctive work of alignment required to hold collaborative time – and the forms of collective activity it underpins – together.

In particular, we seek to account for the role of *non*-human forces and actors in the shaping of time. We fear that in 'socializing' time we may run the risk of *denatur*ing (and even, rather oddly, *dematerializing*) it; or more precisely, obscuring its specific and consequential nature(s) and materialities behind a too-general abstraction. We argue that there remain highly specific categories of time (articulated in the typology that follows) that tend to disappear behind the too-neat distinction between subjective and objective time. In the cases of distributed scientific practice we study, these intersect in a fluid and dynamic way with what might be called the 'social' properties of time, but which here we further articulate as institutional, biographical and infrastructural time. The collaborative rhythms we study are both highly 'natural' and 'material' as well as highly 'social'.

3. MAPPING COLLABORATIVE RHYTHMS

All forms of collective activity, human and otherwise, are subject to rhythm. Things emerge, grow, evolve, and give way to new phenomena according to distinctive patterns. In this paper, we consider those elements of rhythm that touch, impinge on, emerge from, or otherwise implicate the world(s) of distributed collective practice, with principal examples from efforts to organize and design supporting technologies for large-scale collaborative science. In this context, we note three general features of rhythm that cut across each of the more specific typologies offered below.

First, all rhythms are *specific*, emerging from discrete sources and structured according to particular patterns; this sets them apart from the more formalized and abstract categories of time used to mark and track them. Second, as encountered in the real world (as opposed to our neatened analytic descriptions of same), all rhythms are *multiple*, showing up in messy and heterogeneous form and rarely if ever alone. Any given site or activity, or any isolated moment in time, may be best thought of as a gateway or constriction through which multiple rhythms are flowing at once, some of which will be contradictory or dissonant in nature.

Third, all rhythms (at least of the sort we're interested in) are potentially *meaningful*, caught up in the world of perception, interpretation, and experience. This opens up certain representational or 'imaginary' dimensions of time as "real in their effects" – for example, as organizational actors account for and reconstruct rhythms both forwards and backwards (consider here the (contested!) role of origin stories, life histories, and futures in orienting individual and joint action). In the worlds we study, in many regards that matter, the 'experience' of time is inseparable, both practically and analytically, from its 'fact'. Our last and by now hopefully obvious point has to do with the endogeneity of rhythms and the forms of collective action they support. Rhythms are constitutive of distributed collective practices, and vice versa. The whole is the sum of its flows.

If we believe these points to be true in a general sense, they're especially salient in the worlds of scientific collaboration we study. In particular, successful scientific collaborations must seek to accommodate and align four separate kinds or modalities of time, each of which shape and structure the rhythms of collaborative work in specific and often challenging ways:

phenomenal rhythms –the distinctive forms of time emanating from the field and objects of study themselves.. For instance, in the ecological field sciences, these rhythms may be seasonal: animals mate, snow falls and melts, and vegetation grows, buds, matures, and declines according to distinctive. In such cases collaborative work time is organized in part around the phenomena under study. Other rhythms may be more episodic or event-driven in character: in the medical world, medical teams group and pace themselves around the rate of tumor growth, and epidemiologists organize their work practices in part with an eye to the spread rate of diseases. Rare but unpredictable events such as cosmic ray bursts, supernovae, tsunamis or earthquakes require rapid mobilization of teams and equipment.. Other rhythms may be circadian in nature – for example, the patterns imposed by the nocturnal activities of certain species, or the traditionally night time art of astronomy. Still others impose rhythms of a far more extended or truncated sort – for example, efforts to study long-term climate change, or conversely, the splitting of sub-atomic particles. In these and many other fields, phenomenal rhythms carry deep, immediate and often challenging implications for the nature and organization of collaborative work.

institutional rhythms – a second set of rhythms can be found embedded in the organizations and institutions, large and small, that structure and govern scientific work. These range from the rhythms set by local academic calendars (e.g.,the timing of summer and winter breaks, annual patterns marking the arrival and departure of new students and research assistants), to the rhythms established by the deadlines and review processes of national funding bodies, to the (discipline-specific) submission and event dates for key academic conferences. Other institutional rhythms may operate at the lab or research group level – for example the perpetual difficulty of scheduling meetings and joint calls between colleagues balancing radically different teaching and service schedules in their home departments or research units. Like phenomenal rhythms, institutional rhythms may pose collaborative challenges of their own – for example, the difficulties of working with colleagues at institutions with different academic calendars (whether the distinction between 'quarter' and 'semester' systems in the U.S. or the more radical seasonal offset that separates researchers in the northern and southern hemispheres).

biographical rhythms – other temporal patterns and limits emanate from the life choices and circumstances of collaborative participants. This is an often overlooked category of rhythm, largely because it tends to spill across the line between professional and personal lives., . In this category we see the timing of children, illness and recovery, divorces and new relationships, births and deaths. We also see patterns of activity associated with various stages or moments in the development of biographical trajectories, from the doctoral apprenticeship through the pressures of junior faculty development to the post-tenure gravy train, along with rhythms emanating from a variety of less canonical routes (e.g., movements into and out of administration, or back and forth across the lines separating academic from government, industrial, and other locations). Shifting roles, identities, and career trajectories are central constituents of biographical rhythm – though we would note that careers themselves are built (and sometimes challenged) at the intersection of institutional and biographical time.

infrastructural rhythms (or rhythms of the built environment) – a final category of rhythm emanates from the nature and rhythms of the built world itself, including (in our case) the extensive assemblage of equipment and infrastructure attending the production and sharing of scientific knowledge itself. This is the timeliness of machines, artifacts and systems, from the durability of the Periodic Table of the Elements to the development and operation of the Large Hadron Collider. It's the time of software upgrades, hardware replacement schedules, and the time it takes to build adoption of a new protocol, instrument or data standard within a research group or across a field (weighed against the time required to build interoperability between otherwise 'local' systems down the road). It's the time it takes to a spacecraft to Mars and the window of opportunity before the Rovers go dead. In many of the fields we study, the built environment itself imposes certain and often exacting constraints on the nature and rhythm of scientific practice and collaborative work. Large-scale histories of technology have articulated such principles largely as matters of direction, pointing to forms of 'path dependency' that often accompany the development of new technological systems and infrastructures; we argue here that they are also matters of rhythm and pace.

4. HYBRIDS, TENSIONS, AND ALIGNMENT

While the above typology points to collaborative rhythms in their separate and purified forms, temporality in the real world(s) of scientific collaboration and other collective practice rarely shows up in anything like as neat or seamless a form. In practice, collaborative scientific practices combine elements of most, and usually all, of the above. The distinctive temporalities attending specific instances of collaborative work are usually shaped precisely at the intersection of often-contradictory tendencies embedded within and between each of the categories noted above. This makes rhythmic disjuncture or dissonance a frequent and under-examined tension within distributed scientific forms – and the complex art of rhythmic alignment a much-understudied category of organizational work.

Some such tensions have already been hinted at within the category descriptions given above: the alignment challenges posed by different institutional calendars; the tensions attending choices between short-and long-term costs and payoffs in infrastructural development; etc. Such tensions only multiply as we (as analysts) or they (as collaborative participants) move between the categories. What happens when work moves across the purely conceptual lines distinguishing phenomenal, institutional, biographical, and infrastructural time (or more precisely, where the temporal patterns embedded in each fail to mesh)? Our fieldwork suggests that the world of collaborative science is in fact rather full of such mismatches, and just as many efforts (small and large, local and systemic) to ameliorate, deal, or simply live with them. We illustrate such tensions with the following set of stories:

Story 1: Studying Long Term Phenomena on Short-Term Funding

Our first story illustrates a classic tension between phenomenal and institutional time. As academic researchers well know, science has long been funded in short-term chunks, structured in the U.S. around the canonical three-year grant (or shorter still). This poses no particular problems for fields built around discrete experiments – the psychological experiment, the biological lab study, the one-off opinion survey, etc. But what if your phenomenon of study and the methods it requires unfolds on a different sort of timescale (decadal, centennial, millennial, etc.)? For analysts of long-term ecological change, institutional rhythms have long posed a particular challenge. As one ecologist explains,

> Trees grow for hundreds of years, hurricanes may decimate a site every 50 years, and droughts may last for decades; thus, a long-term perspective is needed to understand the ecological response to these slow changes or rare events. (Hobbie 2003).

Such misalignments between short-term process and long-term phenomena have led to some famous and costly errors. For example, the 1922 allocations of water under the Colorado River Compact were based on a period in the early twentieth century that turns out (we now believe) to have been among the wettest in centuries. This has led to the famous problem of 'paper water' in the Southwestern United States (Jackson 2005).

The contemporary Long-Term Ecological Research (LTER) Network has emerged as an effort to redress this misalignment between phenomenal and institutional rhythms. Rather than short term grants LTER is reviewed on a decadal basis, and its 26 geographically distributed sites are reviewed every six years. In this manner LTER has itself become a relatively stable institution for ecological research. At the level of the science this has meant longitudinal monitoring of research sites with an emphasis on data curation, sharing and dissemination. Thus, while the majority of research in ecology is still grant supported, behind these cycles of funding stands an organization oriented to the study of ecological phenomena.

Story 2: Living on Mars Time

Our second story is drawn from the NASA Mars expedition rover (MER) project (as recounted in recent dissertation work by Zara Mirmalek) (Mirmalek 2008). Here the rhythms (and tensions) are multiple, with collaborative activity pulled between the competing demands of phenomenal, institutional, and biographical time. The story begins with a minor (but consequential!) solar discrepancy: the Martian day is precisely 2.7% longer than that on earth. To make up the difference, and to not lose crucial sunlight needed to recharge the Rover's solar batteries, NASA made the decision to put its Rover team on Mars time for the duration of the project. Members of the project team were to live, literally, on Mars time, organizing their work (and broader lives) around a day that was 24 hours and 39 minutes long. Clocks and wristwatches were redesigned to operate on Mars time. As the mission went on, members of the MER team literally drifted across the earth day, as the Martian sunrise moved from morning, to afternoon, to evening, and back again.

As the project progressed, strains between this phenomenally structured time and the normal biographical rhythms of the project team began to emerge. The medical team working with the project noted marked physical consequences for the work team, who began manifesting symptoms that looked like (and amounted to) an interplanetary form of jetlag. Such physical problems were joined by even more pronounced consequences for the personal lives of project participants, who found themselves arriving home to sleeping spouses and children one week, and at breakfast the next. As time passed, many participants opted to essentially live at the lab with their temporally aligned colleagues, rather than face a forever-receding schedule back on Earth.

CONCLUSION

The brief stories offered above suggest just some of the ways in which, in distributed collective practices in the sciences and elsewhere, "rhythm matters." This paper has sought to provide an initial account, theoretical and empirical, for the under-recognized temporal rhythms and challenges that structure collaborative scientific practice – a point meriting further research within organization science, CSCW, and other information school fields. In particular, we argue for the salience of four central and often imperfectly aligned categories in establishing the consequential rhythms of collaborative life: phenomenal, institutional, biographical, and infrastructural. Our present work seeks to build on this understanding, developing new methods, tools, and heuristics for the understanding of collaborative rhythm across a range of distributed collective practices in the sciences and elsewhere.

5. REFERENCES

Bardram, J. E. (2000). "Temporal Coordination: On Time and Coordination of Collaborative Activities at a Surgical Department." Computer Supported Cooperative Work **9**: 157-187.

Barley, S. (1988). On Technology, Time, and Social Order: Technically Induced Change in the Temporal Organization of Radiological Work. Making Time: Ethnographies of High Technology Organizations. F. A. Dubinskas. Philadelphia, PA, Temple University Press**:** 123-169.

Braudel, F. (1992). The Identity of France. London, Perennial Press.

Braudel, F. (2004). Memory and the Mediterranean. London, Vintage.

Clark, H. H. and S. E. Brennan (1991). Grounding in Communication. Perspectives on Socially Shared Cognition. L. Resnick, J. M. Levine and S. D. Teasley. Washington, DC, American Psychological Association**:** 127-149.

Cummings, J. and S. Kiesler (2007). "Coordination costs and project outcomes in multi-university collaborations." Research Policy **36**(10): 1620-1634.

Edwards, P., S. J. Jackson, et al. (2007). "Understanding Infrastructure: Dynamics, Tension, and Design." Report from "History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures".

Giddens, A. (1991). Consequences of Modernity. Palo Alto, CA, Stanford University Press.

Harvey, D. (1991). The Condition of Postmodernity. London, Wiley-Blackwell.

Hinds, P. and M. Mortenson (2005). "Understanding conflict in geographically distributed teams: the moderating effects of shared identity, shared context, and spontaneous communication." Organization Science **16**: 290-307.

Hollan, J. and S. Stornetta (1992). "Beyond being there." Proceedings of the SIGCHI conference on Human factors in computing systems, Monterey, California, USA: 119-125.

Jackson, S. J. (2005). Building the Virtual River: Numbers, Models, and the Politics of Water in California. Department of Communication. San Diego, University of California, San Diego. **Ph.D.**

Jackson, S. J., P. N. Edwards, et al. (2007). "Understanding Infrastructure: History, Heuristics, and Cyberinfrastructure Policy." First Monday **12**(6).

Kiesler, S. and J. Cummings (2002). What do we know about proximity and distance in work groups? A legacy of research. Distributed Work. P. Hinds and S. Kiesler. Cambridge, MA, MIT Press**:** 57-80.

Lakoff, G. and M. Johson (1980). Metaphors We Live By. Chicago, University of Chicago Press.

Lefebvre, H. (2004). Rhythmanalysis: Space, Time, and Everyday Life. London, Continuum.

Lemke, J. (2000). "Across the Scales of Time Artifacts, Activities and Meanings in Ecosocial Systems." Mind, Culture, and Activity **7**(4): 273-290.

Mirmalek, Z. (2008). Solar Discrepancies: Mars Exploration and the Curious Problem of Inter-Planetary Time. Department of Communication. San Diego, University of California, San Diego. **Ph.D.**

Nowotny, H. (1992). "Time and social theory: Towards a social theory of time." Time and Society **1**(3): 421-454.

Olson, G. M. and J. S. Olson (2000). "Distance Matters." Human-Computer Interaction **15**(2-3): 139-179.

Orlikowsi, W. and J. Yates (2002). "It's about Time: Temporal Structuring in Organizations." Organization Science **13**(6): 684-700.

Reddy, M., P. Dourish, et al. (2006). "Temporality in Medical Work: Time also Matters." Computer Supported Cooperative Work **15**: 29-53.

Ribes, D. (2006). Universal Informatics: Building Cyberinfrastructure, Interoperating the Geosciences. Department of Sociology (Science Studies). San Diego, University of California. **Unpublished Ph.D. Dissertation**.

Ribes, D. and G. C. Bowker (2008). Organizing for Multidisciplinary Collaboration: The Case of the Geosciences Network. Scientific Collaboration on the Internet. G. M. Olson, J. S. Olson and A. Zimmerman. Cambridge, MIT Press.

Ribes, D. and T. A. Finholt (2007). "Tensions across the Scales: Planning Infrastructure for the Long-Term." Proceedings of the

2007 international ACM SIGGROUP conference on Supporting Group Work, Sanibel Island, Florida, USA: 229-238.

Schmidt, K. (2002). "The Problem with "Awareness"." Computer Supported Cooperative Work **11**(3): 285-298.

Virilio, P. (1986). Speed and Politics. New York, Columbia University Press.

Going Green with IT: A Study of Energy Consumption by Home and School Information Technology Systems in the College of Information at the University of North Texas

Gerald Knezek
Rhonda Christensen
Tandra Tyler-Wood
Okyoung Lim
William E. Neaville
University of North Texas
Denton, Texas USA
01-940-565-4195
gknezek@gmail.com

ABSTRACT

This paper addresses the strategies introduced by the College of Information at the University of North Texas to monitor and begin implementing approaches to enable the college to move toward the university's strategic goal of becoming a climate-neutral, "Green University" [3]. Data based on monitoring selected office and home inventories of information technology equipment were used to generate estimated production-level use and standby (vampire) consumption of electrical energy. Sample-based estimations and projections regarding 2008 to 2009 progress indicate that that the College of Information at the University of North Texas has made extensive progress over one year in reducing the energy consumed by its information technology systems. New computer and printer systems have greater processing power and capabilities but consume no more power than the ones they replace, when in full operation. The computer systems consume one-half to one-third (and in some cases one-seventh) the power of the previous systems when in 'sleep', 'hibernate' or 'shut down' states that qualify for the standard definition of consuming stand by power. The authors estimate that these actions have brought the college at least halfway toward its five-year goal of becoming climate neutral, during the first year of the initiative.

Subject Descriptors

IT infrastructure, sustainability

General Terms

Management, measurement

Keywords

Energy consumption, standby power, going green

1. INTRODUCTION

One area of interest featured in the iConference Call for Participation is *IT infrastructure development and sustainability in the home, organizations, communities, and society*. This area is consistent with the five-year strategic plan of the University of North Texas (UNT) to become a "Green University": "In April, 2008, UNT became the first major Texas public university to sign the American College & University Presidents Climate Commitment with an ultimate goal of becoming climate neutral" [3]. The College of Information within the University of North Texas receives support from the administration of the university for creating a sustainable office/home infrastructure for IT.

During the summer and fall of 2009, the College of Information (COI) at UNT began to develop and implement approaches to enable the college (and by implication, the entire university) to make significant strides toward the strategic goal. Data based on monitoring selected office and home inventories of information technology (IT) equipment are multiplied by known numbers of systems to generate estimated production-level use and standby (vampire) consumption of electrical energy. Projections are made regarding the progress that has been made over the 2008 to 2009 calendar year toward the "Green University" goal, and how much additional progress is still needed. Since energy monitoring equipment and protocols already exist within the college to support a National Science Foundation (NSF) funded project involving middle school students and their teachers, applications to university campus and home IT systems supporting the university mission have been straightforward. This paper focuses on reducing consumption in "intended use" applications as well as reducing electricity consumption in standby (vampire, wasted power) mode.

2. SIGNIFICANCE/RATIONALE

Global warming has resulted in an increase of about 1.3 degrees Fahrenheit in the annual average temperature of the earth between the beginning and end of the 20th Century [4]. The Intergovernmental Panel on Climate Change (IPCC) has concluded that most of the observed temperature increase since the middle of the 20th century was caused by increasing concentrations of greenhouse gases resulting from human activity such as fossil fuel burning and deforestation [4]. The global surface temperature will probably rise an additional 2.0 to 11.5 °F

during the twenty-first century if major steps to reduce greenhouse gas emissions are not put in place [4]. This will cause sea levels to rise, will expand subtropical deserts [5] will accelerate the retreat of glaciers, permafrost and sea ice; will increase the intensity of extreme weather events, will accelerate species extinctions, and will change agricultural yields.

Most scholars agree that electricity produced by burning coal is the most detrimental producer of greenhouse gas [1], because coal-fired plants produce 2.095 pounds of CO2 for each kilowatt hour of electricity used by consumers [2]. Currently coal accounts for 57 percent of the electricity produced in the United States, so reducing the amount of electricity used is a straightforward means of reducing greenhouse gas emissions [6]. Many universities have taken on the task of reducing their own consumption of electricity as part of the American College & University Presidents Climate Commitment. The University of North Texas is one of those Universities. The College of Information at UNT seeks to be a role model for other colleges in the Going Green initiative at UNT.

3. ABOUT COI AT UNT

The College of Information at UNT has approximately 28 full-time faculty members spanning two departments. Three university research centers and several externally funded projects are also housed within the unit located in the former Texas Instruments manufacturing plant purchased by the University of North Texas and renamed Discovery Park. Much of the information technology equipment initially used by the college was brought by the two departments from their former homes in the Information Sciences and College of Education buildings on the university's main campus. However, significant funds were also made available for new purchases for the new facility.

4. FOUNDATION FOR STUDY IN NSF PROJECT: MIDDLE SCHOOLERS OUT TO SAVE THE WORLD

Beginning in October 2008, researchers in the Institute for the Integration of Technology into Teaching and Learning (IITTL), one of the three centers within the COI at UNT, received funding from the National Science Foundation to train middle school teachers to guide their students in monitoring standby power consumption (power consumed while no useful function is taking place) in household devices such as flat screen TVs, interactive game consoles, and portable power supplies for laptops and other information technology devices. Published estimates ranged from 5% to 20% [8] regarding the portion of electricity in the USA being wasted due to standby power load on the electrical grid, but no comprehensive study has been completed since 1999 [7]. Handheld monitoring devices for 14 classrooms in four US states were purchased, protocols were developed, and classroom teachers were brought together for training in June 2009. As teachers began using "Kill-A-Watt" and "Watts Up?" portable monitoring devices with their students during the fall of 2009, the suggestion was made at a COI faculty retreat that the same equipment and protocols could be applied to office and home IT devices, in order to monitor COI progress toward the UNT President's goal of becoming a Green University within five years of putting the new strategic plan for the university in place in 2008. Thus the current project was spawned.

5. COI PROGRESS DURING YEAR 1 OF THE FIVE YEAR PLAN

Many IT systems brought to the new college from their former faculty and staff locations for 2008-09 were replaced during the fall of 2009. Thus, an initial estimate of progress toward the goal of obtaining "Green University" status could be obtained by measuring normal "in use" energy consumed by old and new systems, as well as standby power consumed when the machines sat idle. These values could then be multiplied by the numbers of systems in use and the hours typically used in each state, resulting in an absolute estimate of kilowatt hours consumed as well as proportional reduction in load on the electrical grid, from old to new installations. This two-department, college-based estimate could then be extrapolated to the 58 departments in the 35,000 student university to obtain an estimate of the progress that could be made by moving to new, more energy efficient devices through the university. In a second phase of the study, recommendations regarding new energy saving devices and techniques with high potential return (such as unplugging all photocopiers during the 4 weeks the university is closed during Christmas break) could be produced. For phase 1, selected initial estimations of savings are listed below.

5.1 Office Computer Systems

Microsoft Windows operating system-based units are the officially supported information technology workstations at the University of North Texas, and therefore most systems used by faculty and staff are of this type. As of fall 2008, Windows tower computers manufactured by UNT's Microcomputer Maintenance Shop with a 15-17 inch CRT monitor, formed the typical workstation. As of 2009, most systems have been replaced by Dell Optiplex 950 computers with Dell AX510PA 17 inch flat panel LCD monitors.

As shown in Table 1, the 90 new Windows systems in the College of Information at UNT consume on average about one-fourth to one-third less power when in full use or screen-saver mode, versus the systems they replaced, and two-thirds to three-fourths less power when put to sleep or shut down. The Dell system power consumption when shut down (1.9 Watts) can be further decomposed into 1.2 watts for the computer and .7 watts for the attached monitor, which means both devices are near to complying with the latest Energy Star guidelines of "less than 1 watt" of standby power consumed, and in fact the monitor does comply. However, if the more typical definition of "being put to sleep", which allows "instant on" is taken as the standby power criterion, then the monitor still complies by virtue of consuming 1.0 watt of the 3.7 shown, but the computer itself still has a ways to go as it consumes 2.7 watts when put to sleep. The 90 new units are estimated to save more than 1150 pounds of CO2 not put into the atmosphere each year by the extra power that would have had to be produced by the coal-burning plants that provide most of the Dallas-Ft. Worth Metroplex electricity (at 2.095 pounds of C02 per kilowatt hour). The cost savings to the university is approximately $555 per year in reduced electrical cost, for 4625 fewer kilowatt hours used, based on an estimated nationwide current average price of $.12 per kilowatt hour for electricity.

Table 1. Comparison of 2008 versus 2009 Typical Windows Office Workstation in College of Information at UNT.

System	In Use	Screen Saver	Sleep Mode	Shut Down	Total
2008 Windows + CRT Monitor	117.2	165	9.3	8.2	
2009 Windows + Flat Panel LCD Monitor	91.7	92.5	3.7	1.9	
Difference	25.5	72.5	5.6	6.3	
Hrs. per Day	6.0	2.0	4.0	12.0	
Days per wk	5.0	5.0	5.0	2.0	
Wks per yr	30.0	30.0	30.0	22.0	
KWH/Year	23.0	21.8	3.4	3.3	51.5
$/KWH	0.1	0.1	0.1	0.1	0.1
$/Unit/Yr	2.8	2.6	0.4	0.4	6.2
COI Units	90.0	90.0	90.0	90.0	90.0
Cum KWH/yr	2065.5	1957.5	302.4	299.4	4624.8
Cum $/year					$555.0
LB CO2/Yr					1162.7

In addition to the Windows workstations listed in Table 1, during 2009 the college purchased nine 21-inch flat panel iMac computers that are all-in-one units. These were placed in offices and labs in main departmental areas and are considered equal in processing power to the Dell Optiplex computers previously described. Each of these uses an average of 113.5 watts when turned on and in use, just slightly less than the Dell/Windows workstations listed in Table 1 (117.2 watts). However in screensaver mode the iMac uses considerably less (110 watts vs. 165 for Dell), while in sleep mode (1.7 watts) and when shut down (1.0 watts) they use much less than the 9.3 and 8.2 watts for the Dell systems. The iMacs appear to comply with the newest Energy Star guidelines for standby power in electronic appliances (< 1 watt in standby mode) and could be one avenue pursued by the University to reach its Going Green goals. Note that these machines typically replaced 15-inch eMac self-contained Apple computers that used built-in CRT screens. These eMac machines consumed 91.4 watts when in use, 7.7 watts when in sleep mode and 3.2 watts when shut down. As shown in Table 2, when in sleep mode or shut down, the iMacs use only 1/3 the power of the machines they replaced. The new iMacs consume roughly one-half the power of their Dell peers in the sleep or shutdown modes.

Table 2: Comparison of Self-Contained eMac Computers versus iMac Computers in College of Information at UNT

System	In Use	Sleep Mode	Shut Down
eMac	91.4 watts	7.7 watts	3.2 watts
iMac	117.2 watts	1.7 watts	1.0 watts

Color printers were also upgraded in the UNT College of Information during 2009. As of fall, 2009, the college housed eight, workgroup class, networked printers. These Dell Phaser 8860 color laser printers use approximately 162 watts when in operation and 11.5 watts in sleep (power saver) mode. This usage is much less than the 860 watts in full operation and 31.6 watts consumed in power saver mode by the HP Color Laser 4600 printers many of the Dell printers replaced. The new laser printers use only 1/3 as much electricity as the old.

5.2 Home-Based IT Systems

Ten faculty and staff in the College of Information at UNT took the "Watts Up?" energy monitoring devices home during September – October 2009 in order to provide data on their home IT systems used to support work functions. The number of reported plug-in appliances per home varied from two to twelve, with a computer and printer forming the minimal home configuration. Selected examples of old versus new systems for home computers and printers will be supplied for the purposes of providing an estimate of progress underway (and possible in the future) for IT used in the home.

One faculty member reported upgrading a computer and color printer during the 2008-09 time frame. A 15-inch Apple Emac (91.4 watts in operation, 7.7 watts in sleep mode, 3.2 watts shut down) was replaced by a 19-inch flat panel iMac (93.5 watts in use, 0 watts in sleep mode, 0 watts shut down). An HP 4430 Color Laser Printer (860 watts in operation, 31.6 watts in power saver, 0.2 watts switched off) was replaced by a Dell 2135 Color Laser Printer (90.5 watts in operation, 14.5 watts in hibernate mode, 0 watts switched off). The old computer used three to seven times more power than the new in sleep/shut down mode, while the old printer used twice as much power in power saver/hibernate mode. One unexpected revelation was that apparently all new Apple laptops with break-away power supply cables power down to 0 watts used in sleep mode (and shut down mode) after the battery is charged. Some newer Windows laptops were found to have this feature as well, such as a Toshiba 15" laptop tested by one home user. Apparently the power "bricks" themselves are smart enough to power down when the current drain drops to a small amount of the initial load.

5.3 One Area with Little Progress

Many IT workstations within the College of Information at UNT have universal power supply (UPS) battery backup systems to help prevent loss of data in the event of power loss. The standard unit provided at UNT consumes approximately 20.5 watts continuously even after the battery is charged, and even if the workstation is unplugged. Certainly hundreds of these must exist among the computers used by the 1830 faculty / staff on campus.

A UPS that powers itself down, similar to the power bricks on newer laptops, would make a great step forward toward energy conservation in this area.

6. SUMMARY/CONCLUSIONS

Data gathered during the summer and fall of 2009 indicate that the College of Information at the University of North Texas has made great progress from 2008 to 2009 in reducing the energy consumed by its information technology systems. New computer and printer systems have greater processing power and capabilities but consume no more power than the ones they replace, when in full operation. They consume one-half to one-third (and in some cases one-seventh) the power of the previous systems when in 'sleep' or 'hibernate' or 'shut down' states that qualify for the standard definition of consuming stand by power. For some devices, such as universal power supply battery backup systems, little progress has been made. However, overall we can conclude that the types of replacement equipment purchased in the College of Information over the past year, if applied through the entire university, would bring UNT at least halfway toward its five year goal of being a "Green University" in the area of IT systems.

7. ACKNOWLEDGMENTS

Our thanks to university faculty and staff for providing energy consumption data.

8. REFERENCES

[1] Borenstein, S. 2007. *Carbon-emissions culprit? Coal.* The Seattle Times Company. DOI= http://seattletimes.nwsource.com/html/nationworld/2003732690_carbon03.html

[2] Department of Energy. 2000. *Carbon Dioxide Emissions from the Generation of Electric Power in the United States*. DOI= http://tonto.eia.doe.gov/FTPROOT/environment/co2emiss00.pdf.

[3] Himmel, J. 2009. University of North Texas. DOI= http://www.wimba.com/company/events/dls/#dls-archived

[4] IPCC (2007-05-04). *Summary for Policymakers* (PDF). Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment Report of the Intergovernmental Panel on Climate Change. DOI= http://ipcc-wg1.ucar.edu/wg1/Report/AR4WG1_Print_SPM.pdf

[5] Lu, J., Vecchi, G. A.; Reichler, T. 2007. Expansion of the Hadley cell under global warming. *Geophysical Research Letters* **34**: L06805. doi:10.1029/2006GL028443. http://www.atmos.berkeley.edu/~jchiang/Class/Spr07/Geog257/Week10/Lu_Hadley06.pdf.

[6] Power Scorecard. 2000. *Electricity from: Coal*. Pace University. DOI= http://www.powerscorecard.org/tech_detail.cfm?resource_id=2. White Plains, NY.

[7] Rosen, K. B., & Meier A. K. 2000. *National energy use of set-top boxes and telephony products*. Berkeley, CA: Lawrence Berkeley National Laboratory.

[8] Ross, J.P., & Meier, A. 2000. In Proceedings of the Second International Conference on Energy Efficiency in Household Appliances, Naples, Italy.

[9] The American College & University Presidents' Climate Commitment. 2008. *Climate Leadership for America: Progress and Opportunities in Addressing the Defining Challenge of Our Time*. DOI= http://www2.presidentsclimatecommitment.org/reporting/documents/ACUPCC_AnnRep_2008.pdf

A Sense of Wonder: Enhancing Access to Folktales through Task and Facet Analysis

Kathryn La Barre
University of Illinois
501 East Daniel, MC-493
Champaign, IL 61820
001.217.244.4449
klabarre@illinois.edu

Carol L. Tilley
University of Illinois
501 East Daniel, MC-493
Champaign, IL 61820
001.217.265.8105
ctilley@illinois.edu

ABSTRACT
Discusses the approach taken in Phase 1 of a three-phase project Folktales, Facets and FRBR [funded by a grant from OCLC/ALISE]. This project works with the special collection of folktales at the Center for Children's Books (CCB) at the University of Illinois at Urbana-Champaign, and the scholars who use this collection. The project aims to enhance the effectiveness and efficiency of folktale access through deep understanding of user needs. Phase 1 included facet analysis of the bibliographic records for a sample of 100 folktale books in the CCB, and task analysis of interviews with four CCB-affiliated faculty. Describes the information tasks, information seeking obstacles, and desired features for a discovery and access tool related to folktales for this initial group of scholarly users of folktales.

Categories and Subject Descriptors
H.3.1 [**Content Analysis and Indexing**]: Indexing Methods. H.3.3 [**Information Search and Retrieval**]: Search Process.

General Terms
Performance, Design, Theory.

Keywords
Task analysis, Facet analysis, Search and Discovery.

1. INTRODUCTION
Folktales connect communities and people across time and space with each story evolving in its transformation from performance to text. Even as stories change, they continue to carry culturally unique values and ideals, anchored in shared human experience, with continuing relevance for its audience. Beyond their use in the specific communities where they are born, folktales find audiences among people of all ages and educational levels—from children hearing about stone soup for the first time to an established scholar examining the transmission of treasure tales in the Dominican Republic.

A network of informants, adapters, compilers, storytellers, librarians, and scholars keeps stories alive through telling, collecting, and publishing, yet these efforts are frequently undermined by existing structures for representation and discovery in the bibliographic catalogs of libraries and similar institutions, potentially obscuring these stories from continued study and use. For instance, the records for single volume collections of tales seldom provide complete or searchable information about the titles and origins of each story in the volume, thus requiring a searcher to intuit a book's potential relevance and persevere to undertake an examination of the physical volume.

New strategies are needed in order to overcome the shortcomings of information retrieval systems that may hamper efficient information seeking for complex information resources such as folktales. The development of new strategies is complicated specifically for folktales because of the heterogeneity of users and tasks. For instance, scholarly users may want to undertake a comparative study of a particular tale type, while librarians designing a children's program may want to find multiple versions of a single tale in order to identify the most appropriate one for their needs (Goldberg, 2003). Even children, on their own or with the assistance of their caregivers, might want to explore different retellings of a favorite story such as East of the Sun and West of the Moon. Yet, each of these users must often rely on the brief descriptions in bibliographic records as they attempt to complete their information tasks.

Through an iterative combination of facet and task analysis that supports deep understanding of information tasks and allows the creation of new access models, this project aims to enhance discovery of and access to folktales and related resources. Kuhlthau's (2005) call for greater connection between the study of users' information-seeking behaviors and the design of information retrieval systems to better enable collaborative frameworks which would encourage and strengthen task-focused information seeking studies and user-centered system design motivates this project. The research design also draws from recommendations for more integrated and theoretically repositioned models for information seeking and use and

information retrieval (e.g. Hjørland, 1997; Ingwersen & Jarvelin, 2005).

2. METHOD AND FINDINGS

The scope of this preliminary work is limited in three important ways. First, the researchers interviewed only a small number of subjects, each of whom is engaged in scholarly activity related to folktales but none of whom would be identified primarily as folklorists. Second, the folktale collection that forms the basis for the facet analysis is comprised largely, but not wholly, of folktales that have been adapted for a juvenile audience; some folklorists (e.g. Goldberg, 2003) would consider a collection such as this one inadequate to support legitimate folklore scholarship. Third, as the informants and researchers are colleagues, the possibility for bias in both response and interpretation is amplified.

2.1 Task analysis

Task analysis is a repertoire of techniques commonly used in the field of human-computer interaction to support the development of systems and interfaces from a user-based perspective. In the past decade, task analysis has increasingly been used to understand people's information seeking processes (Vakkari, 2003). No universal definition of tasks exists and it can be difficult to disambiguate task and goal, but for the purposes of this study, task is best defined as any information-seeking activity necessary to complete some scholarly goal (cf. Xie, 2008). A first step in conducting any task analysis, then, is to understand users' goals.

As part of this preliminary study, we conducted one-hour semi-structured interviews with four of the five faculty members who use this collection, in order to determine the manner in which they conduct research in this area as well as their use of the CCB collection. Although not conventionally folklorists, the subjects are engaged in scholarly activity related to folklore, including editing collections of folktales, reviewing folktales adapted for children, studying audience engagement in storytelling performance, and documenting the history of literary transmission of folktales. Each of the subjects also teaches in the area of youth services librarianship, which has a strong tradition of oral storytelling (cf. Hearne, 1998), so folklore permeates their discussions and work with students. Finally each of the subjects has performed folktales orally as part of professional work experiences outside academe.

The purpose for these interviews was to ascertain 1) folktale-related scholarly practices (i.e. goals); 2) obstacles the informants have encountered in information seeking; and, 3) suggestions for an ideal tool that would help them in their information-related activities. Although we asked direct questions to elicit relevant insights, we also asked each subject to talk more broadly about other areas including their experiences working with folktales and their educational experiences related to folktales in order to capture information relevant to our interests that may not have been revealed through direct questioning. The interviews were recorded and transcribed for coding, after which we developed the coding framework on an emergent and iterative basis in order to identify scholarly practices.

2.1.1. Scholarly practices in folklore

Six categories of scholarly practices surfaced in the interviews:

(1) *Exploring* (e.g. Reading tale collections for possible future uses; monitoring websites or journals to stay current on scholarly issues pertaining to folktales)

(2) *Creating* (e.g. Adapting a folktale for performance; designing a library program based on a folktale)

(3) *Synthesizing* (e.g. Critiquing a published adaptation of a folktale for a juvenile audience; documenting the published variants of a particular tale; preparing lecture notes and other instructional materials)

(4) *Studying* (e.g. Conducting research on audiences' responses to oral performance; examining the relationship between women's personal narratives and folktales)

(5) *Collecting* (e.g. Building a personal folktale library to support scholarship; keeping notes about folktale variants to support scholarship)

(6) *Searching* (e.g. Using a bibliographic tool to identify a variant; following cited references to identify relevant information)

Some of the goals overlap with Palmer, et al's (2009) synthetic model of scholarly information practices. For instance, she and her co-authors identified "collecting" as a core scholarly activity. Searching appears in their model as well as ours, but we have a related category—"exploring"—as well that represents non-directed searching activities that we identified; in contrast, Palmer, et al subsume a similar activity—"browsing"—beneath "searching." The activity Palmer, et al termed "writing" is similar to "synthesizing" that emerged from our data. Both "studying" and "creating" are unique to our framework with the latter category representing an activity similar to "synthesizing" but with a greater emphasis on creative transformation.

We anticipate conducting further interviews and observation of practice with these scholars and other folklorists, as part of the next phase of our research. From this added observation and interview data, we will be able to refine the list of tasks and to identify the tasks essential to supporting successful goal completion for these users.

2.1.2. Obstacles to information seeking

Regarding obstacles to information seeking, the interview data clustered in two categories: disciplinary-related and discovery and access-specific. Examples from the latter category are not especially unique in that they relate to lack of awareness of useful bibliographic tools or problems with the tools themselves (e.g. quickly outdated).

More interesting are the disciplinary-related obstacles, several of which touch on the variable nature of folktales. For instance, the names given to tales may vary from one collection or one community to another; similarly, tale variants may share motifs, although the variants have quite different effects or themes. A disciplinary-related obstacle such as this one, however, also speaks to problems with existing tools (e.g. limited or no cross-references).

Another intriguing set of disciplinary-related obstacles pertains to "translating," or working across boundaries (cf. Palmer, et al, 2009). The subjects identified translation problems as they sought and accessed information from a variety of scholarly (e.g. literary criticism, psychoanalysis, anthropology) and disciplinary perspectives (e.g. structuralist, historical-geographic). Translation problems also occurred as subjects moved from understandings of tales informed by personal experiences (e.g. recalling stories told by family members, reading tales in childhood) to understandings constructed through scholarly practice.

2.1.3. Desired features for search and discovery tools

The features these subjects identified as essential for an ideal discovery and access tool for folktale scholarship reflected both their work as scholars and their professional experiences in storytelling and youth services librarianship. For instance, the scholarly focus is evident in requests for searchable fields for source notes and cultural attributions, as well as descriptor fields for motifs such as characters. The professional focus is clearly visible in proposing the inclusion of programming ideas, ties to learning standards, and suggested audience ages for performance. All subjects indicated preferences for a tool that would permit both directed searching and serendipitous discovery and that offered extended synopses or the full text of tales for searching.

Worth noting is that some existing bibliographic record structures such as MARC already partially support informants' requests (e.g. for a searchable field for cultural attribution), although these structures are seldom leveraged fully to these ends. In a separate paper (Tilley and La Barre, 2010) we offer a provisional model for bibliographic records that shows where existing MARC fields overlay those arising from this study.

2.2 Facet analysis

This study proposes that facet analysis is a necessary and useful first step towards the creation of user-oriented search and discovery systems. The facet-analytic dimension of this study builds on a traditional understanding of facets, as articulated by Ranganathan, who viewed them as basic concepts that are inherent in a given subject. A facet may be a concept, characteristic, attribute or aspect that may assist in the identification of a set of distinct entities. Facets are uncovered through a technique known as facet analysis, which requires the conceptual analysis of a subject area into a set of fundamental categories. The essence of facet analysis is the sorting of terms in a given field of knowledge into homogeneous, mutually exclusive facets, each derived from the parent universe by a single characteristic of division (Vickery, 1960 p. 12). The entire process of facet analysis is governed by a canon composed of principles (specific rules), postulates (guidelines) and devices (Vickery, 1960). Additional guidance for the facet-analytical approach used in this study comes from Cochrane (1965).

After establishing a complete shelflist of folktale books in the CCB's collection, we created a stratified (i.e. by decade of publication) random sample of 100 folktale books to form the core collection of materials that were subjected to facet analysis in Phase 1 of this project. The CCB is one of the world's premier reviewing and examination centers for children's books and related materials. It houses more than 15,000 English-language print and non-print resources in its non-circulating collection. Folktales published in single-tale volumes and multiple-tale collections, and scholarly resources related to folklore and storytelling comprise approximately ten percent (or 1500 items) of the collection. Publication dates for the print materials span the 20th and 21st centuries, but a majority of the items were published after 1960; this distribution is reflected in our sample.

For each item in our sample, we examined several artifacts for the facet-analysis portion of the protocol. First, we examined the books themselves. Second, we inspected the local bibliographic records as well as the most complete bibliographic records for each item we were able to obtain through WorldCat. Finally, we scanned reviews—primarily those published in the *Bulletin of the Center for Children's Books* but also from other sources—for items in the sample. The hope is that this will present an opportunity to uncover a variety of facets that might be useful beyond those typically represented by the fields now being leveraged in library catalogs.

Our analysis of facets is still ongoing at this preliminary phase. For instance, we have yet to engage in a rigorous facet analysis of users' information tasks, or deep facet analysis of the indices and controlled vocabularies used by folklore scholars to assist them in locating relevant stories. The names given to each facet may not be entirely reflective of the terms used by scholars. Both the terms used, and the facet groupings will be subject to further refinement as more interviews and observations are conducted and subject to facet analysis. Refinements are also expected as a result of further facet analysis of the subject access tools, such as folklore-specific controlled vocabularies and classifications used by folklorists.

2.2.1 Preliminary facets derived from the collection

Based on the preliminary analysis, we have identified the following facets (*in italics*) and areas where the focus of each facet may be sharpened or refined [in parentheses]:

Agent [may include: author/narrator, translator, adapter, editor/compiler, illustrator, etc.]

Area [of source] [of story]

Association [award] [aggregations of multiple stories] [related materials] [stylistic dependencies] [source] [work]

Content [characters] [illustrations] [language] [mood] [moral] [motif] [narrative structure] [story type]

Context [age of story] [audience] [function of story] [language of source] [manner of dissemination] [style] [type of variant]

Documentation [external sources like bibliographies or indexes]

Genre [type of story]

Origin [cultural] [ethnic] [geographical] [theoretical] [of source]

Time [of source] [of story]

Transmission [oral] [print] [function]

Viewpoint [theoretical] [cultural] [ethnic]

As the list indicates, a variety of facets emerged through our analysis. For instance, folklore publications are typically careful to articulate authorial responsibility, or *agent*, by clearly distinguishing among authors, editors, adapters, translators, illustrators, and retellers. Another important facet pertains to *genre*, which acknowledges differences among story types such as: folktales, fairy tales, fables, legends, and myths. The *origin* facet indicates cultural attribution—whether according to geographic region or by reference to a particular ethnic or cultural group. The *documentation* facet supports cross references or direct linkage to external sources such as notes or bibliographies. Such linkage is especially important for recently published works which may be available in digital, full text format. [Motif], here shown as a focus of the *content* facet, emerged as another important characteristic of folklore material. Within folkloric analyses of folktales, motifs refer to small persistent elements of individual stories such as actors (e.g. a princess, Baba Yaga), items (e.g. a magical stick, a curse), and plot elements (e.g. a contest, burial alive)(cf. Thompson, 1946).

These preliminary facets echo several aspects of Uther's recommendations to guide the creation of new motif indexes and related tools to assist folklorists. "In establishing concepts for new indexes and integrating the narrative material for a region or ethnic group, the following should be required:"

(1) clearly defined time and area,

(2) theme-oriented presentation,

(3) indication of structural elements

(4) chronological and structural listings of variants,

(5) suggestions of related items,

(6) year of publication,

(7) references to external sources and literature,

(8) indexing by subject, names, places, narrators (Uther, 1997, p. 215).

2.2.2 Facets derived from bibliographic tools

In addition to examining the sample of books from the CCB, we also examined selected bibliographic tools to aid in the discovery of and access to folktales (e.g. Ashliman, 1987; MacDonald and Sturm, 2003; American Folklore Society, n.d.) along with some core scholarly and overview works related to folktales (e.g. Dorson, 1972; Thompson, 1946; Toelken, 1996). Suggestions for the works we examined came both from our interview subjects and from bibliographies such as the one provided by the Folk Narrative Section of the American Folklore Society. Our analysis of these materials provided further support for the validity of the facets derived from the book sample.

Many, but not all, of the preliminary facets already have underlying bibliographic record fields that may support facet display, but are not fully leveraged by library catalogs. For instance, La Barre (2010) queried the term "folktale" in 200 library catalogs using one of six next-generation integrated library systems (ILS)(e.g. AquaBrowser, Koha) in order to determine which facets are currently used and supported. She found the following facets (number in parentheses refers to the number of ILS systems using each facet):

- *subject/topic (6)*
- *author (5)*
- *date of publication (5)*
- *format (4)*
- *genre (4)*
- *location (4)*
- *availability (3)*
- *language (3)*
- *series (3)*
- *call number (2)*
- *subject: geography (2)*
- *subject: time (2)*

3. IMPLICATIONS AND CONCLUSION

Full-text resources have become ubiquitous, whether through subscription databases or digitization projects or the Internet. Add to this reality the perception held by many laypersons and scholars (and even some librarians) that a few keyword searches performed in Google will retrieve a universe of information. The result is that too often the value of providing systematic, reliable, and meaningful access to the intellectual contents of texts is negated. Libraries themselves play a role in this negation when, in an effort to save human and financial resources, they increasingly rely on copy cataloging, purchased records, and other similarly conceived records to provide access to the resources in their collections, believing that these frequently all too minimal descriptions will provide adequate access.

Yet, the ability to search full-text sources is not the *automagical* tool some scholars and laypersons would have us believe. In a study still relevant today, Blair and Maron (1985) demonstrated that the recall rate for relevant documents when users used free-text searching in a large data set not constructed for the purpose of testing retrieval was on average below 20%.

As Blair and Maron argued, "it is impossibly difficult for users to predict the exact words and combinations, and phrases that are used by *all* (or most) relevant documents and *only* (or primarily) by those documents" (295). Even in an era with improved natural language processing algorithms to facilitate searching, the results are unsatisfactory (e.g. Tomlinson et. al 2007).

Folktales are but one example of resources that are often obscured by the movement to full-text searching and the subsequent reduction in the provision of rich bibliographic records. Oral histories, archival materials, museum artifacts, musical scores, and many other types of texts are similarly difficult to for users to locate. By seeking to understand how users of these resources integrate them into their work tasks, and by systematically analyzing the domains in which these resources are situated, we look to design alternative models for bibliographic records that highlight, rather than obscure, these resources for the people who turn to them most frequently.

4. ACKNOWLEDGMENTS

Support for this preliminary research was provided by a generous grant from OCLC and the Association of Library and Information Science Education. Carrie Pirrman also provided valuable research assistance; she is presenting a poster at iConference 2010 that highlights corollary research.

5. REFERENCES

[1] American Folklore Society. Ethnographic Thesaurus Online. <et.afsnet.org>.

[2] Ashilman, D. 1987. A guide to folktales in the English language: Based on the Aarne-Thompson Classification System. Greenwood Press.

[3] Blair, D. C., Maron, M.E. 1985. An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System. Communications of the ACM 28 (3): 289-299.

[4] Cochrane, P. 1965. American Institute of Physics, Documentation Research Project: A review of work completed and in progress, 1961-1965. American Institute of Physics.

[5] Dorson, R. 1972. Folklore and folklife: An Introduction. U of Chicago.

[6] Goldberg, C. 2003, Folktale research and the Pantheon Fairy Tale and Folklore Library. J Am Folklore. 116 (Spring 2003) 217 – 218.

[7] Hearne, B., 1998, Midwife, witch, and woman-child: Metaphor for a matriarchal profession. In Story, from fireplace to cyberspace : connecting children and narrative (Papers presented at the Allerton Park Institute held October 26-28, 1997), ed. Betsy Hearne, et al. 37 – 51

[8] Hjørland, B., 1997, Information seeking and subject representation: An Activity-Theoretical Approach to Information Science, Greenwood.

[9] Ingwersen, P., Jarvelin, K. 2005. The Turn: Integration of information seeking and retrieval in context, Springer-Verlag.

[10] Kuhlthau, C. C., 2005, Towards collaboration between information seeking and information retrieval, *Inform Res* 10 (2) paper 225.

[11] La Barre, K. 2010. Facets, Search, and Discovery in Next Generation Catalogs: Informing the Future by Revisiting Past Understanding. In *Paradigms and conceptual systems in Knowledge Organization, Proceedings of the eleventh international conference of the International Society of Knowledge Organization,* ed. Claudio Gnoli.

[12] MacDonald, M.R., Sturm, B. 2003. The Storyteller's sourcebook: A subject, title, and motif index to folklore collections for children, 1983 – 1999. Gale.

[13] Palmer, C., Teffeau, L., Pirmann, C., 2009. Scholarly information practices in the online environment: Themes from the literature and implications for library service development, www.oclc.org/programs/publications/reports/2009-02.pdf

[14] Thompson, S. 1946. *The Folktale*. Dryden Press.

[15] Tilley, C., La Barre, K. 2010. New Models from Old Tools: Leveraging an Understanding of Information Tasks and Subject Domain to Support Enhanced Discovery and Access to Folktales. In *Paradigms and conceptual systems in Knowledge Organization, Proceedings of the eleventh international conference of the International Society of Knowledge Organization,* ed. Claudio Gnoli.

[16] Toelken, B. 1996. The dynamics of folklore (Rev. and Exp. Edition). Utah State U.

[17] Tomlinson, S., Oard, D.W., Baron, J.R., Thompson, P. 2007. Overview of the TREC 2007 Legal Trac. In *The Sixteenth Text Retrieval Conference (TREC 2007) Proceedings* (NIST Special Publication SP 500-274)

[18] Uther, H. 1997. Indexing folktales: A critical survey. J Folklore Res 34 (3): 209-220.

[19] Vakkari, P. 2003. Task-based information searching. ARIST. 37. 413-464.

[20] Vickery, B. C. 1960. Faceted classification. A guide to the construction and use of special schemes. ASLIB.

[21] Xie, I. 2008. Interactive information retrieval in digital environments. IGI.

Service Science in iSchools

Kelly Lyons
Faculty of Information
University of Toronto
Toronto, ON, Canada
+1-416-946-3839
kelly.lyons@utoronto.ca

ABSTRACT

In this paper, we argue that the discipline of service science, in search of an academic home, is ideally situated within an iSchool curriculum and research community. We describe the features and expectations of the emerging field of service science and the skills identified as important for service scientists, and compare them to the features and expectations of iSchools along with skills important for information professionals. We present results of a systematic review of iSchool universities by identifying which faculties, schools or departments in those universities are pursuing courses, programs, and research in service science and demonstrate that most of the effort today is taking place from within business schools or engineering departments. We then discuss the opportunities, impact, and benefits of situating service science research and education programs within iSchools, thereby arguing why iSchools provide an ideal environment for the study of service science.

Keywords

Service science, iSchools, impact, research, education

1. INTRODUCTION

Gathering momentum in the early 1990's the iSchool movement has grown to include 24 schools at 23 universities in six countries in 2009 [9, 16]. The core vision of the iSchools involves bringing a multidisciplinary approach to the study of information, technology, and people as equally interacting entities [16]. A motivation for the emergence of iSchools is a tremendous growth in the amount of digital information [10].

On a slightly later timeline (early 2000's), a call to action or movement was taking place in a topic area being called *service science* (and, in some cases, *service science, management, and engineering*) [2]. Service science is the study of *service systems* which vary in scope (from individuals to businesses, organizations, governments, and nations) and involve people, information, organizations, and technology adapting dynamically and connecting internally and externally to other service systems through value propositions [14]. Service science strives to bring together many disciplines (computer science, systems engineering, cognitive science, economics, organizational behavior, human resources management, marketing, operations research, and others) in an attempt to study and understand service systems [14]. A motivation for the emergence of service science is the fact that the service sector is the fastest growing in most economies yet it lacks strong conceptual foundations [2].

Companies, led by IBM, began turning to academics to help determine the foundational concepts necessary for a science of services to emerge [2]. Programs and courses began appearing in universities around the world from within different departments and schools. Some institutions were building service science programs in an attempt to bring more excitement to computer science which was experiencing a decline in enrollment [13]. Yet many service science programs are growing out of business schools in marketing, management, operations research, and administration.

In this paper, we look at the 23 universities that have iSchools in the iSchool Caucus (iCaucus) [9] and determine which faculties, programs and/or schools in those universities have service science research and teaching activities. In most cases, business schools have been the place from which service science has emerged. In several of the universities analyzed, engineering schools house service science programs. Only five of the 23 universities in the iCaucus have service science programs, activities, or courses within their iSchool.

We argue that there is a tremendous opportunity to bring impact both to iSchools and to the service science community by engaging in service science research and education through an information and iSchool lens. We formulate this argument by 1) comparing the goals and needs of service science initiatives to those of the iSchool movement; 2) describing and comparing the skills and knowledge of service scientists with those of information professionals; and, 3) describing one program within an existing, yet new, iSchool curriculum.

2. SERVICE SCIENCE PROGRAMS

We did an analysis of the 23 universities with iSchools listed as part of the iSchool caucus to determine which universities have a department or school that has identified service science or service science, management, and engineering as a program, center, course, or have individuals participating in research in service

science. Our research was carried out using the following method:

- Each iSchool website was accessed and courses, programs, research projects, and faculty member profiles were reviewed looking for references to service science, service science, management or engineering (SSME), or related topics
- Each university web site was searched for information about service science, management or engineering, or related topics
- A web search was conducted on the 23 university names with each of the terms "service science", "services science", "SSME", and "service, science, management, and engineering"
- Conference programs from key service science conferences [3, 4, 5] were analyzed for references to the 23 universities

We categorized service science participation as: *programs* which include degrees, certificates, or combinations of courses in service science; *courses* which include one or more individual courses offered on service science; or, *activities* which include individual faculty members or students participating in workshops or conferences on service science or individual research programs in service science. We found that, of the 23 universities with an iSchool, only three have no service science program or activities, twelve have some form of service science involvement through their business or marketing schools, eight have some in their engineering or computer science departments, and only five have a service science presence in their iSchool (some universities have service science presence in more than one discipline). Table 1 summarizes the results of our findings by identifying those universities and their school(s) or department(s) with *programs*, *courses*, or *activities* that are explicitly described as service science efforts. Most of the service science activities in universities with iSchools have emerged from the business schools and their established programs in service marketing or service operations. It is not clear that a discipline of service science can effectively evolve from business school programs (or engineering schools, for that matter) when there are many other facets to service science and a requirement to bring together many disciplines (computer science, systems engineering, cognitive science, economics, organizational behavior, human resources management, marketing, operations research, and others) [7].

Table 1: List of iSchool universities and service science activities (if any) at that university outside of and within the iSchool.

University and Corresponding iSchool		Service Science?		
		Program	**Course**	**Activities**
University of California, Berkeley	School of Information	(Engineering; CS; iSchool ✓; other disciplines) [7]	(Engineering; CS; iSchool ✓; other disciplines) [7]	(Engineering; CS; iSchool ✓; other disciplines) [7]
University of California, Irvine	The Donald Bren School of Information and Computer Sciences	--	--	--
University of California, Los Angeles	Graduate School of Education and Information Studies	--	--	Management
Carnegie Mellon University	School of Information Systems and Management, Heinz College	Computer Science	Computer Science	Computer Science
Drexel University	College of Information Science and Technology	--	--	Business
Florida State University	College of Communication and Information	--	--	Business
Georgia Institute of Technology	College of Computing	--	--	Engineering
Humboldt-Universitat zu Berlin	Berlin School of Library and Information Science	Not available	Not available	Not available
University of Illinois	Graduate School of Library and Information Science	--	--	Business Administration
Indiana University	School of Informatics and Computing	--	--	Business
Indiana University	School of Library and Information Science	--	--	Business
University of Maryland	College of Information Studies	Business	--	Marketing
University of Michigan	The School of Information	--	--	iSchool ✓
University of North Carolina	School of Information and Library Science	--	--	Marketing

The Pennsylvania State University	College of Information Sciences and Technology	--	--	iSchool✓; Engineering; Hospitality and Management
University of Pittsburgh	School of Information Sciences	--	--	Business
Royal School of Library and Information Science, Denmark		--	--	--
Rutgers, the State University of New Jersey	School of Communication and Information	--	--	Engineering
Singapore Management University	School of Information Systems	iSchool✓	iSchool✓	iSchool✓
Syracuse University	School of Information Studies	--	--	--
University of Texas, Austin	School of Information	--	--	Marketing; Business; Science, Technology and Society
University of Toronto	Faculty of Information	--	iSchool✓	iSchool✓
University of Washington	Information School	Engineering	--	--
Wuhan University, China	School of Information Management	Software School	Software School	Software School

Some iSchools have realized the similarities in goals with service science and are working to build programs in service science. In describing their service science, management, and engineering program, the Dean of the Singapore Management University iSchool said, "Over the past five years, our School of Information Systems (SIS) has been pioneering new educational approaches right at the intersection of information technology, IT solutions and business needs – which means we have been working in the spirit of service science all along." [15]

While most iSchools do not have direct link to or work in the emerging field of service science, many have courses and programs with similar goals and content. For example, the Masters of Information Management degree program at the University of Maryland iSchool describes teaching future information professionals "... what they need to understand to manage issues related to users of information, the organization, the content, the technology, and the global environment without being experts in each one of them." [16]. This capability is similar to that of service scientists who need to have knowledge in human, business, and technical issues and should be deep problem solvers in one or more disciplines but have the ability to capably interact with and understand specialists and concepts from other disciplines [8]. In the next section, a more detailed comparison of the motivations and challenges of iSchools and those of service science is presented which further highlights the similarities between the two.

3. SERVICE SCIENCE AND iSCHOOLS

According to Maglio and Spohrer [14]: "Service science combines organization and human understanding with business and technological understanding to categorize and explain the many types of service systems that exist as well as how service systems interact and evolve to cocreate value." Service science, therefore, brings knowledge and understanding in organizations, humans, business, and technology to bear in a study of service systems which are made up of people, information, organizations, and technology. We compare that to iSchools which, "are interested in the relationship between information, people and technology" [9]. Given this comparison, one can see that the scholarly pursuits within iSchools are well suited to the study of service systems.

The iSchools' vision addresses the multidisciplinary requirements necessary for progress to be made in science, business, education, and culture: "An iSchool provides the venue that enables scholars from a variety of contributing disciplines to leverage their individual insights, perspectives, and interests, informed by a rich, 'trans-disciplinary' community." [12]. Service science requires bringing together understanding and knowledge in different areas (organizational and human, business, technology) and different disciplines together to study service systems [14]. Crossing academic disciplines is difficult for a number of reasons including the fact that each discipline has different methods, norms, values, goals, models, ethics, and ways of interacting with external bodies [11]; however, these multidisciplinary hurdles are not only being overcome in iSchools but being exploited to enhance the research and education of information professionals [11]. The multidisciplinary focus in iSchools makes them an ideal place to engage in service science research and teaching.

Another component of the iSchool vision is the importance of industry which can help shape an applied research agenda and also the leadership iSchools bring in providing direction to industry and government. An important strategy in the creation of a science of service is engaging with university, government, and industry partners [19]. The importance of conducting research that is defined jointly with industry and academia is well-understood in iSchools and in the field of service science.

3.1 Education and Knowledge

In addition to similarities in motivations and challenges with respect to content and vision, iSchools and service science have related knowledge needs and ideals for graduates of their programs. The iSchools' vision states that expertise in all forms of information is needed to make progress and must include "understanding of the uses and users of information, as well as information technologies and their applications." [11]. Service

science is the study of interconnections of people, technology, organizations, and information that requires knowledge and understanding in human and organizations, business domains, and technology [14]. There are significant relationships between the kinds of knowledge and expertise needed to study all forms of information and that needed to study the entities that make up service systems.

Some people have stated that progress in service science requires people who are "T-Shaped" [7, 8] with deep knowledge (the vertical part of the "T") in one or more areas but able to bridge and communicate across the complexities of the other disciplines (the horizontal part of the "T") – others have described "π-Shaped" people which better symbolizes the requirement for depth in more than one area. A similar concept is described by the Georgia Institute of Technology College of Computing iSchool in a document outlining their undergraduate computer science program. It states that their program helps students become symphonic-thinking people who "will develop expertise in multiple, high-value areas of computing and act as innovative boundary crossers." [6] This perspective is motivated by several recent writings including that of Pink in which he identifies that: "The last few decades have belonged to a certain kind of person with a certain kind of mind – computer programmers who could crank out code, lawyers who could craft contracts, MBAs who could crunch numbers. But the keys to the kingdom are changing hands. The future belongs to a very different kind of person with a very different kind of mind –creators and empathizers, pattern recognizers, and meaning makers." [18].

It is no longer sufficient for people to have expertise in one area without understanding the connections and contexts of that expertise in relationship to other disciplines. This is especially true in a multidisciplinary field such as service science and in iSchools where information, technology, and people are of roughly equally significance [17] such that knowledge and understanding in one (no matter how deep that knowledge reaches) is not sufficient without connections to the others.

Table 2 summarizes the similarities identified between the iSchool vision and the goals and motivations of the emerging discipline of service science.

Table 2. Summary of service science and iSchool motivations

iSchool Vision	Service Science Goal
Interested in relationship between information, people, and technology	Studies service systems which are made up of information, people, technology, and organizations
Requires multidisciplinary approach	Requires multidisciplinary approach
Work with industry to shape research direction	Bring academia, industry, government together
Must bring understanding in uses and users of information, information technologies, applications	Must bring understanding in human and organizations, technology, and business domains
Depth in one of information, technology, people not sufficient to understand connections between them	Requires depth in one or more areas and the ability to communicate across complexities of other disciplines

3.2 Service Science iSchool Program

In this section, we describe the Universit of Toronto iSchool program with several specializations including one in architectures and services which includes a service science course, and service science research activities. We show how the program connects to and fits within the existing iSchool program and discuss some of the benefits of participating in a service science research and education program within an iSchool.

The iSchool@Toronto offers a masters of information program with four core courses in: 1) knowledge and information in society; 2) representation, classification, organization and meaning-making; 3) information systems, services and design; and, 4) an information workshop that integrates concepts across the core courses using a set of team-based, hands-on activities. Students then choose a path of study: critical information studies; library and information science; archives and records management; information systems, media and design; cultural heritage; knowledge management and information management; or, the more general field of information. Within each path, clusters of courses can make up a further specialization. In the field of information systems, media and design, students may specialize in services and architectures which includes a course on service science.

This introductory course is available to masters level students across the areas of specialty within the iSchool. It was originally taught in a computer science graduate program. Variants of that original course have been offered twice since in the iSchool. The course is broken into four main topic areas each covered in approximately one quarter of the term:

1. Introduction to service science and service systems: What is it? Why is it important? How does the service system lens apply (or not) in information service contexts?
2. Modeling, analyzing, and optimizing service systems: review of literature and hands-on experience
3. Innovation in service systems: What new technologies, work practices, or business models can be used to enhance service systems?
4. Service oriented architectures (SOA) and service oriented computing: How can SOA be used to implement and architect service systems?

The first assignment asks students to select an organization, entity, infrastructure, or institution to analyze as a service system. They analyze their selected system according to various criteria and definitions presented in the service science literature. We have found that most of the service science concepts and definitions, coming from business schools, have been applied in business and government instances of service systems. In the iSchool, students test the published service science concepts on different kinds of service systems such as cultural and public institutions, non-profit organizations, and information-focused entities. By applying an information or iSchool lens to service science concepts, we have been able to challenge notions such as those of *customer* and *recipients of value* as they are currently defined in service science literature. This has led to many interesting discussions, analyses, and new research topics and directions. Our experience provides an example of how iSchools can bring impact and important diverse perspectives to studies in service science.

4. CONCLUSION

In this paper, we compared the motivations and challenges of service science with that of the iSchools. We argue that there is considerable benefit to studying service science from within iSchools. These benefits include being able to explicitly investigate information as a significant part of service systems, having access to multiple disciplines and perspectives within one school or faculty, looking critically at the technological, domain, and social aspects of service science and their connections and interactions, and studying service science concepts with relation to social good, public institutions, and benefit to society.

It is not yet clear whether service science will grow out from an existing academic field or emerge as an entirely new discipline [7]. It is clear, however, that a multidisciplinary approach is needed and it can be much easier to forge ties and build programs within one academic discipline such as an iSchool than across the existing academic silos that are faculties, schools, departments, and colleges.

Despite the similarity in service science and iSchool challenges and motivations, there are very few iSchools engaged explicitly in service science research or teaching. There are, however, a variety of service science programs that have been launched from other departments and schools within the universities that house the iSchools. There is an opportunity to bring service science knowledge, teaching, and research activities within the scope of an iSchool. Not only will the iSchools be able to branch into and help define an important emerging discipline [1, 20] and have access to industry partnerships and funding in the area, but service science will benefit tremendously from the iSchools' multidisciplinary approach to research and knowledge about the connections among information, technology, and people.

5. ACKNOWLEDGMENTS

The author thanks the reviewers for feedback that improved the presentation of this paper and her colleagues at IBM and in the University of Toronto iSchool for valuable discussions and their continued interest in and support of her Service Science pursuits.

6. REFERENCES

[1] America Competes Act, Public Law 110–69—Aug. 9, 2007, Retrieved November 17, 2009 from http://arpa-e.energy.gov/public/pl110-69.pdf

[2] Chesbrough, H. and Spohrer, J. 2006. A research manifesto for services science. Communications of the ACM, Vol. 49, No. 7 (July 2006), ACM Press, New York, NY, 35-40.

[3] Frontiers in Service. 2007. 16th Annual Frontiers in Service Conference Schedule, Retrieved November 18, 2009 from: http://fvortal.cimerr.net/ssme/entry/16th-Annual-Frontiers-in-Service-Conference-Conference-Schedule

[4] Frontiers in Service. 2008. 17th Annual Frontiers in Service Conference, Retrieved November 18, 2009 from: http://www.rhsmith.umd.edu/frontiers2008/

[5] Frontiers in Service. 2009. 18th Annual Frontiers in Service Conference, Retrieved November 18, 2009 from: http://www.rhsmith.umd.edu/ces/eBrochure.html

[6] Furst, M. and DeMillo, R. A. 2006. Creating symphonic-thinking computer science graduates for an increasingly competitive global environment, Whitepaper, Retrieved November 18, 2009 from: http://www.cc.gatech.edu/education/undergrad/bscs/threads_whitepaper.pdf

[7] Glushko, R. J. 2008. Designing a service science discipline with discipline. IBM Systems Journal, 47(1), 15-27.

[8] IfM and IBM. 2008. Succeeding through service innovation: A service perspective for education, research, business and government. Cambridge, United Kingdom: University of Cambridge Institute for Manufacturing. ISBN:978-1-902546-65-0. iSchools, Retrieved November 18, 2009 from: http://www.ischools.org/

[9] iSchools Motivation, Retrieved November 18, 2009 from: http://www.ischools.org/history/motivation/

[10] iSchools Organization, Retrieved November 18, 2009 from: http://www.ischools.org/history/organization/

[11] iSchools Vision, Retrieved November 18, 2009 from: http://www.ischools.org/history/vision/

[12] Jones, E. L., Owensby, J. N., and Allen, C.S. 2008. Strategy for inserting SSME into the undergraduate experience at a minority serving institution. In Service Science: Research and Innovations in the Service Economy, Service Science, Management and Engineering Education for the 21st Century, Eds., B. Hefley and W. Murphy. Springer, NY, 141-146.

[13] Maglio, P. and Spohrer, J. 2008. Fundamentals of service science. J. Academy of Marketing Science, 36(1), 18-20.

[14] Miller, S., Dean of the School of Information Systems, Singapore Management University, Service Science, Management and Engineering (SSME) Programme, Retrieved November 18, 2009 from: http://www.sis.smu.edu.sg/programme/SSME/

[15] MIM Program, University of Maryland, Retrieved November 18, 2009 from http://ischool.umd.edu/programs/mim.shtml

[16] Olson G. M. and Grudin, J. 2009. The information school phenomenon. In Interactions, 16, No. 2 (March and April 2009). ACM Press, New York, NY, 15-19. DOI=http://doi.acm.org/10.1145/1487632.1487636

[17] Pink, D. H. 2005. A whole new mind: Moving from the information age to the conceptual age. Riverhead Books, New York.

[18] Spohrer, J. 2009. Welcome to our declaration of interdependence, Service Science 1(1), i-ii, 2009, Retrieved November 18, 2009 from http://www.sersci.com/ServiceScience//upload/12273996030.pdf

[19] Spohrer, J. and Riecken, D. Guest Editors, 2006. Communications of the ACM, July 2006, 49(7), 31-32

[20] Wobbrock, J. O., Ko, A. J., and Kientz, J. A. 2009. Reflections on the future of iSchools from inspired junior faculty. In Interactions, ACM Press, New York, NY, 69-71. DOI=http://doi.acm.org/10.1145/1572626.1572641

Theory and Education: A Case of Structuration Theory

Lai Ma
School of Library and Information Science
Indiana University – Bloomington
1320 E 10th Street, LI 011
Bloomington, IN 47405-3907
lama@indiana.edu

ABSTRACT

The everyday work of information professionals is culturally, socially, and organizationally structured. The education of information professionals should not be limited to the teachings of practical skills and should include theoretical knowledge. A theory-based education provides the necessary intellectual tool for information professionals for critiquing, evaluating, improving, and refining practices, on the one hand, and for reflecting on the authority, legitimacy and acceptability of professional standards and policies and their cultural, ethical, and social implications, on the other. This paper suggests that structuration theory can make a valuable contribution to iSchool education, for it provides the necessary concepts for the study of the interrelationship between the everyday work of information professionals and "the social" that provoke critical and reflective thinking.

Categories and Subject Descriptors

K.3.2 [**Computer and Information Science Education**]: Information systems education

General Terms

Theory

Keywords

Structuration theory, critical social theory, education

1. INTRODUCTION

The work of information professionals is situated in social space and is culturally and socially structured. Database design, web programming and the uses of different kinds of information technologies are all social activities. These activities usually have as a goal improving the design and development of information services, systems, and policies; at the same time, they also serve certain organizational, economic, and political functions. The everyday work of information professionals is not the mere performance of certain skills but involves the understanding of the "the social" as well as the cultural, ethical, and social implications of their actions and activities. This paper suggests that structuration theory can make a valuable contribution in iSchool education, particularly the concepts that make possible the study of the interrelationship between everyday work and "the social" that provoke critical and reflective thinking.

1.1 "The Social"

In 1887, the first library school in the United States was named the "School of Library Economy" by Melvil Dewey. The term "social epistemology" first proposed by Margaret Egan and Jesse Shera in the early 1950s, emphasizing the role of library in social change, has been an ongoing topic of interest within the library and information science community [15, 16, 18, 20]. In 1968, Shera published a book entitled Sociological Foundations of Librarianship. In the 1990s, Rob Kling described the research area "social informatics" for the study of the relationship between information technologies and social life [26]. In the recent decade, scholarly works in the domains of the social studies of science, the sociology of science, and science and technology studies (STS) have been influential in research in information science (for example, [6, 27, 28]. Most recently, Cronin has discussed "the sociological turn in information science" [11]. Awareness of "the social" has been present since the inauguration of library school in the United States. However, it seems that little has been said about how it should be studied and how it can be incorporated in education of information professionals, not to mention many discussions of "the social" are not grounded in social theory.

Indeed, despite the many discussions of "the social" in IS research, they have not been broadly incorporated in most IS schools' curricula. Chu's content analysis of the curricula of ALA accredited LIS programs in the United States shows that there are very few "theory-based" courses [9]. Budd's review of management education in library and information science suggests that course materials usually emphasize management skills rather than concepts such as authority and responsibility and do not often address ethical issues [8]. One possible reason for the low number of theory-based courses in LIS may be the lack of recognition of the importance of theories by some practitioners, which is quite apparent in the recent discussion of the ALA Task Force Recommendations on Education [1, 39].

In response to the common misconception that theory is "abstract" and is therefore irrelevant to day-to-day practices, one can argue that it is actually the other way around: theory is an intellectual tool for critiquing, evaluating, improving, and refining practices. For, only if information professionals learn about and understand theories of communication can they evaluate and improve information services such as system design and

construction of indexing terms; and only if they learn about and grasp theories of social systems can they reflect critically on the ways in which their work is affected by and has implications for the organization within which they work and with which they interact such as library and university systems, funding agencies, private sector businesses, and local, state and federal governments. Kling has explicitly argued for "critical professional education" in library and information science, for he had witnessed the failures of information system designs that had cost millions of dollars because of the insufficient understanding of "the relationships of IT configurations, socio-technical interventions, social behavior of other participants in different roles, and the dynamics of organizational and social change" of IT professionals [25, p. 395]. Audunson has also aptly pointed out that a reflective practitioner "is not only taught to repeat established practices, but to go behind them, to criticize, refine and develop these practices and discard them if necessary" [2, p. 104]. Thompson also shares the view that the understanding of the relationship of theory and practice is germane to information studies [38].

Theoretical knowledge is as important as practical knowledge in iSchool education. Indeed, researchers in information science have imported theories from many different disciplines and schools of philosophy [12-13]. It is time for similar importation to take place in iSchool education. In what follows, I will briefly discuss Anthony Giddens' theory of structuration and how it could be useful for information science research and education, particularly Giddens' insights into the relationship between social interaction and social systems.

2. The Theory of Structuration

Anthony Giddens' theory of structuration has been very influential in the social sciences for his analysis of the relationship between structure and agency, in which concepts such as power, identities, contexts, and social systems are discussed. The theory of structuration is a response to the two poles of social theories at the time of Giddens' writing in the 1970s and 80s: the structuralist and functionalist, on the one hand, and the hermeneutic and interpretive, on the other. For Giddens, the structuralist and functionalist view of "societal totality" neglects the importance of human actions in the constitution of society. At the other extreme, hermeneutic and interpretive sociologies view actions and meanings only in terms of human conduct and experience and thus neglect "external" factors such as contexts and constraints. The concept of society and the study of the social sciences, for Giddens, however, is "neither the experience of the individual actor, nor the existence of any form of societal totality, but social practices ordered across space and time" [22, p. 2, emphasis added].

Central to the theory of structuration is the concept of "structure." For Giddens, structure is a virtual order, that is, structure does not have a physical existence. Rather, it is manifested in and through routinized activities involving the applications of rules and the manipulations of resources. Rules include those that are overt and formal (for example, Library Bill of Rights) as well as those that are covert and escape explicit awareness (for example, organizational practices). Resources, on the other hand, are either allocative, involving "command over objects, goods or material phenomena," or authoritative, meaning "command over persons or actors" [22, p. 33]. Giddens illustrates the nature of structure using spoken language as an example: when we utter a sentence in English based on formal grammar and social norms, the utterance is a manifestation of structure and thus demonstrates "structural properties"; at the same time, the utterance reproduces English. At the same time, although speaking English involves the understanding of formal grammar and social norms, at most times we do not invoke and are not aware of these formal and informal rules in our conversations. Structure is thus nowhere to be "seen" but is always involved in social interactions. Giddens posits the "duality of structure" in which structuration is a process depending on structural properties that are "both medium and outcome of the practices they recursively organize" [22, p. 25]. In structuration theory, the concept of structure emphasizes the role of social interactions in the constitution of society, in contrast to the structuralist and functionalist view of "societal totality." The understanding of the interrelationship between structure and action also denies the epistemological assumptions of interpretive sociology that adhere to psychologism. Moreover, the duality of structure is important to Giddens' claim that "all social research has a necessarily cultural, ethnographic or 'anthropological' aspect to it" [22, p. 284] because one cannot explain social phenomena without understanding human agency, social interactions and other "structural properties" and their relations to "context" and social systems.

It should be clear by now that structure is not a physical entity; rather, it is virtual and is always evolving, although the rates of change may vary widely in different parts of society and among societies. Structure is implicated in all social interactions because they are dependent upon rules and resources. The analyses of social interactions and routinized activities, however, must also involve the conceptualizations of human agency, context and their relationships which make structuration possible. For Giddens, humans are "purposive agents" in the sense that they are aware of the intended consequences of their own actions. This awareness is the "reflective monitoring of action" based upon rountinized practices. At the same time, routinized practices provide human agents with "ontological security" because members of a community then understand "how to go on" with their day-to-day activities with the knowledge of possible consequences. Structure is thus implicated within the mutual relationship of human agency and routinized activities. Moreover, this mutual relationship is contextual. "Context" in structuration theory is not mere "background," "environment," or "container; rather, it co-evolves with social interactions. That is to say, "context" not only involves time, space, and local settings, but also the co-presence of other agents which make social interactions and hence maintenance and reproduction of structure possible. Structuration theory thus rejects the structuralist, functionalist, and interpretive views of society and suggests a social ontology understood as coordinated human activities and the conditions of these activities.

3. Structuration Theory and Education

The everyday work of information professionals is culturally, socially, and organizationally structured. The education of information professionals should not be limited to the teachings of practical skills and should include theoretical knowledge such as the concepts of structuration theory. This is not to say that practical skills are not important, but that they should be complemented with theoretical knowledge. It is because theoretical knowledge provides the necessary tool for critical and

reflective thinking about the "why" and "how" of professional practices, standards, and policies such that they can be evaluated, refined, and discarded if necessary. Rosenbaum, one of the first scholars to make use of structuration theory in LIS, has conceptualized "information use environment" for the understanding of the complexities of the work environment of information professionals [34-35], and related issues such as the relationship between agency, technology, and organization. He suggests that the usually system- or user-centered system designs and evaluations are due to the lack of necessary *concepts* for the analyses of social interaction such as those of the theory of structuration can provide. In other words, without theoretical knowledge, information professionals lack the necessary tools—concepts—for the evaluation and improvement of information services and systems.

Orlikowski, a management and information systems scholar, has also introduced structuration theory into information systems research [for example, 29-30, 32, 41]. Her work focuses on how technology and organization are interconnected based on the concept of "duality of structure." With Baroudi, she has also argued for the use of interpretive and critical research methodologies in information systems in response to the limitations of positivistic, "descriptive" research [31].

Since then, there has been growing interest in social theory in information science. Articles citing structuration theory, or its "surrogates" such as works of Orlikowski and Rosenbaum have also been slowly increasing. For example, Bouthillier [5] has applied the concepts of structuration theory for the analysis of meaning of service in a public library. Cronin [10] has discussed the potentials of structuration theory for the study of scholarly communication. In the following I will expand the discussion of major concepts of structuration theory and their potential contributions to information science research and education.

3.1 Duality of Structure, Standards and Policies

The duality of structure is one of the very important concepts in structuration theory. It explains that "structure" is both the medium and outcome of the conduct it recursively organizes and that structural properties do not exist outside of human actions and are implicated in the production and reproduction of social systems. The concept is important for information professionals because it suggests that everyday professional activities are not merely the performances of certain skills or the accomplishments of certain tasks, but also the production and reproduction of "structure," and that these professional activities are influenced by and have implications for professional standards, practices, and policies. For example, while the routinized practice of cataloging in academic libraries is a seemingly mundane activity conforming to explicit rules (for example, Anglo-American Cataloging Rules (AACR)) and to organizational practices and norms, the act of cataloging itself is actually maintaining and reproducing the authority of AACR and the Library of Congress, on the one hand, and the organizational practices and norms, on the other.

Indeed, the rules and procedures involved in the use of programming and mark-up languages in information retrieval systems, the compliance with federal laws and institutional rules in creating user surveys, the institutionalized procedures of collection development and most other routinized activities of information professionals, though usually escape explicit awareness, are the medium as well as the outcome of the social practices enacted within specific cultural and social milieus. These routinized activities continue and reproduce the authority, legitimacy and acceptability of certain professional practices, standards and policies. The realization of the duality of structure not only makes possible a deeper understanding of existing rules and practices, but also enables critical and reflective thinking about the possible intended and unintended consequences of the act of following these rules and practices, and as such, their ethical, cultural, and social implications.

3.2 Action, Structure, and Practicalities

The theory of structuration is also useful in a more practical sense. Rosenbaum [35] has pointed out that the design of information services and systems should neither be system- nor user-centered. Rather, it should be based on the analyses of routinized practices of social interactions. The design of interactive interfaces and information systems, the construction of indexes, and most other professional activities in information science involve the understanding of human-human and human-computer interactions and how their interactions are related to social situations. The concept of structure in structuration theory provides the epistemological foundations for these analyses. For example, system-centered designs often neglect the "structure" of human interactions involved and in turn lead to the "non-usable" designs as described in Forsythe' and Kling's works [for example, 17, 25]. The user-centered approach, on the other hand, is often oriented toward the understanding of "information behaviors," or "information needs," and neglects the relationship between "structure" and larger social systems involved in information seeking activities. Structuration theory is potentially contributive to information system design in that it provides the theoretical foundations for reflecting the dynamics of information system design beyond the system- and user-centered approaches. The understanding of the interrelationship between structure and action will bring us more practical and user-friendly designs and services.

4. CONCLUSION

A theory-based education is necessary for iSchool. Indeed, the first graduate school in library science, the Graduate Library School at the University of Chicago, has championed the idea that a theory-based education is a necessary component of professionalization [37]. Benoît [3-4] has suggested critical theory as a foundation for pragmatic information systems design as well as for developing a critical theoretical perspective in information science. Radford [33] has also suggested the introduction of theory of communication in LIS curricula, particularly for courses such as reference services. Audunson argues that librarians should be taught "epistemology and theory of knowledge in order to be able to critically analyze the epistemological presuppositions of different systems" [2, p. 103]. In this paper, I have discussed the potential contributions of Giddens' theory of structuration for IS research and education. I have also shown that structuration theory is applicable in many professional activities in LIS and, more importantly, it provides the necessary concepts for reflecting and refining practices, standards, and policies. In sum, the understanding of "the social" and its relationship to the work of information professionals are essential for the assessment, evaluation, and improvement of information services and systems.

Structuration theory is one critical social theory that provides the necessary concepts for critical and reflective thinking on the practices, standards, and policies of information professionals.

5. ACKNOWLEDGMENTS

I would like to thank Dr. Howard Rosenbaum for commenting on an early draft of this paper.

6. REFERENCES

[1] Archives of JESSE@listserv.utk.edu. Retrieved July 30, 2009, from http://listserv.utk.edu/archives/jesse.html

[2] Audunson, R. 2007. Library and information science education--Discipline profession, vocation? Journal of Education for Library and Information Science 48(2), 94-107.

[3] Benoît, G. 2001. Critical theory as a foundation for pragmatic information systems design. Information Research 6(2).

[4] Benoît, G. 2002. Toward a critical theoretic perspective in information systems. Library Quarterly 72(4), 441-471.

[5] Bouthillier, F. 2000. The meaning of service: Ambiguities and dilemmas for public library service providers. Library & Information Science Research 22(3), 243-272.

[6] Bowker, G. C., & Star, S. L. 2000. Sorting things out: Classification and its consequences. Cambridge, MA: The MIT Press.

[7] Braman, S. 2006. Change of state: Information, policy, and power. Cambridge, MA: The MIT Press.

[8] Budd, J. M. 2003. Management education for library and information science. Advances in Library Administration and Organization 20, 149-163.

[9] Chu, H. 2006. Curricula of LIS programs in the USA: A content analysis. In C. Khoo, D. Singh & A. S. Chaudhry (Eds.), Proceedings of the Asia-Pacific Conference on Library & Information Education & Practice 2006 (A-LIEP 2006), Singapore, 2-6 April 2006 (pp. 328-337). Singapore: School of Communication & Information, Nanyang Technological University.

[10] Cronin, B. 2005. The hand of science: Academic writing and its rewards. Lanham, MD: Scarecrow.

[11] Cronin, B. 2008. The sociological turn in information science. Journal of Information Science 34(4), 465-475.

[12] Cronin, B., & Meho, L. I. 2008. The shifting balance of intellectual trade in information studies. Journal of the American Society for Information Science and Technology 59(4), 551-564.

[13] Cronin, B., & Meho, L. I. 2009. Receiving the French: a bibliometric snapshot of the impact of 'French theory' on information studies. Journal of Information Science 35(4), 398-413.

[14] Day, R. E. 2007. Kling and the "critical": Social informatics and critical informatics. Journal of the American Society for Information Science and Technology 58(4), 575-582.

[15] Egan, M., & Shera, J. H. 1952. Foundations of a theory of bibliography. Library Quarterly 22(2), 125-137.

[16] Fallis, D. 2006. Social epistemology and information science. Annual Review of Information Science and Technology 40, 475-519.

[17] Forsythe, D. E. 2001. Studying those study us: An anthropologist in the world of artificial intelligence. Stanford, CA: Stanford University Press.

[18] Frohmann, B. 2001. Discourse and documentation: Some implications for pedagogy and research. Journal of Education for Library and Information Science 42(1), 12-26.

[19] Froehlich, T. J. 1989. The foundations of information science in social epistemology. Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences (pp. 306-315). Washington, D.C.: IEEE Computer Science Press.

[20] Furner, J. 2004. 'A brilliant mind': Margaret Egan and social epistemology. *Library Trends 52(*4), 792-809.

[21] Giddens, A. 1979. Central problems in social theory: Action, structure and contradiction in social analysis. Berkeley: University of California Press.

[22] Giddens, A. 1984. The constitution of society: Outline of the theory of structuration. Berkeley, CA: University of California Press.

[23] Golder, S. A., & Huberman, B. A. 2006. Usage patterns of collaborative tagging systems. Journal of Information Science 32(2), 198-208.

[24] Gorman, M. 2003. Whither library education? Paper presented at the Joint EUCLID/ALISE Conference.

[25] Kling, R. 2003. Critical professional education about information and communications technologies and social life. Information Technology & People 16(4), 394-418.

[26] Kling, R. 2007. What is social informatics and why does it matter? The Information Society 23, 205-220.

[27] Latour, B. 1987. Science in action: How to follow scientists and engineers through society. Cambridge, MA: Harvard University Press.

[28] Latour, B., & Wooglar, S. 1986. Laboratory life: The construction of scientific facts. Princeton, NJ: Princeton University Press.

[29] Orlikowski, W. J. 1992. The duality of technology: Rethinking the concept of technology in organizations. Organization Science 3(3), 398-427.

[30] Orlikowski, W. J. 2005. Material works: Exploring the situated entanglement of technological performativity and human agency. Scandinavian Journal of Information Systems 17(1), 183-186.

[31] Orlikowski, W. J., & Baroudi, J. J. 1991. Studying information technology in organizations: Research approaches and assumptions. Information Systems Research 2(1), 1-28.

[32] Orlikowski, W. J., & Yates, J. 2002. It's about time: Temporal structuring in organizations. Organization Science 13(6), 684-700.

[33] Radford, M. L. 2001. Encountering users, encountering images: Communication theory and the library context.

Journal of Education for Library and Information Science 42(1), 27-41.

[34] Rosenbaum, H. 1993. Information use environments and structuration: Towards an integration of Taylor and Giddens. Proceedings of the ASIS Annual Meeting 30, 235-245.

[35] Rosenbaum, H. 1996. Structure and action: Towards a new concept of the information use environment. In S. Hardin (Ed.), Proceedings of the 59th Annual Meeting of the American Society for Information Science, 33. Medford, NJ: Information Today, Inc. 152-157.

[36] Solomon, P. 1997. Discovering information behavior in sense making. I. Time and timing. Journal of the American Society for Information Science 48(12), 1097-1108.

[37] Sugimoto, C. R., Russell, T. G., & Grant, S. 2009. Library and information science doctoral education: The landscape from 1930-2007. Journal of Education for Library and Information Science 50(3), 190-202.

[38] Thompson, K. M. 2009. Remembering Elfreda Chatman: A companion of theory development in library and information science education. Journal of Education for Library and Information Science 50(2), 119-126.

[39] Unsworth, J. 2009. iSchools reply to ALA task force report. Retrieved August 10, 2009, from http://nora.lis.uiuc.edu/images/iConferences/alaresponse.pdf

[40] Yates, J., & Orlikowski, W. J. 1992. Genres of organizational communication: A Structurational approach to studying communication and media. Academy of Management 17(2), 299-326.

Metadata Realities for Cyberinfrastructure: Data Authors as Metadata Creators

Matthew S. Mayernik
Department of Information Studies
Graduate School of Education & Information Studies, UCLA
00+1+3102060029
mattmayernik@ucla.edu

ABSTRACT

Cyberinfrastructure systems for digital data will depend on effective ways of creating and sharing metadata. In distributed scientific collaborations, creating and collecting metadata is a significant challenge. Metadata creation is often an unfunded mandate. We present a preliminary study of metadata creation by data authors in a large science and technology research center. We asked researchers to create metadata using the Dublin Core-based metadata fields for inclusion in a center-wide metadata repository. The results of our pilot test indicate that data authors face a number of challenges in creating metadata, including organizing group vs. individual knowledge, adapting an unfamiliar metadata scheme to the specifics of their project, and drawing boundaries between data sets.

Topics

Information management
Information technology and services
Nature and scope of *i*Schools and *i*Research
Preserving digital information

Keywords

Cyberinfrastructure, metadata, scientific data practices

1. INTRODUCTION

Metadata is a key component of data storage, sharing and preservation systems. Quality metadata helps to facilitate the management, discovery, access, and use of data resources. Cyberinfrastructure systems for digital data will depend on effective ways of creating and sharing metadata [5, 11]. In distributed scientific collaborations, however, creating and collecting metadata is a significant challenge. Metadata creation is often an unfunded mandate. Information or data specialist positions are not yet common in cyberinfrastructure projects [14]. Many cyberinfrastructure projects therefore rely on data authors to create metadata that can be discovered and used by others outside the projects. Little research has examined the experiences of researchers in distributed research projects in creating metadata to be shared in public or community data repositories. Understanding how data authors approach the task of metadata creation – what their understandings of metadata are, what problems they encounter, and their work practices in performing the task – will provide guidance in developing metadata collection policies, processes, and technological systems for future cyberinfrastructure projects.

In this paper, I outline a study of metadata creation by data authors in a large science and technology research center. Researchers within the center are being asked to create metadata for a new center-wide metadata repository as a means to make their research products more visible to the scientific community. We are studying the challenges data authors face when creating new metadata for potential users outside the center using an unfamiliar schema. Preliminary findings indicate that researchers face a number of challenges, including organizing group vs. individual knowledge in creating metadata, adapting an unfamiliar metadata scheme to the specifics of their project, and drawing boundaries between data sets.

2. BACKGROUND

Metadata can be defined in various ways, from "data about data" to "descriptive information about data that explains the measured attributes, their names, units, precision, accuracy, data layout and ideally a great deal more. Most importantly, metadata includes the data lineage that describes how the data was measured, acquired or computed" [9]. Metadata can be created through both automated and manual processes. Both of these methods present challenges. Automated metadata creation techniques exist for text-based documents, but these techniques do not extend to creating metadata for scientific data, as a significant proportion of scientific data is not text-based. Further, automatic techniques require customization for every new type of data creation instrument, as the particulars of the data creation instrumentation and processes are a critical component of metadata descriptions. Much metadata creation thus depends on manual effort.

Additionally, the responsibility for creating metadata falls on different individuals depending on the institutional setting. The National Science Board *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century* report [12], outlines four main actors who play important roles in the data collection and curation process:

- *Data creators*: the scientists, educators, students, and others involved in research that produces digital data.
- *Data managers*: the organizations and data scientists responsible for database operation and maintenance.

- *Data scientists*: the information and computer scientists, database and software engineers and programmers, disciplinary experts, curators and expert annotators, librarians, archivists, and others, who are crucial to the successful management of a digital data collection.
- *Data users*: the larger scientific and education communities, including their representative professional and scientific communities. (pg. 25-28)

Swan and Brown [16] describes how the "data scientist" role, as presented in the Long-Lived Digital Data report, did not accurately portray the roles they observed in a study of data management practices in the United Kingdom research community. They instead identify the important data management roles as the following:

- *Data creators or data authors:* researchers with domain expertise who produce data. These people may have a high level of expertise in handling, manipulating and using data, gained through experience and as a result of need or personal interest…
- *Data scientists:* people who work where the research is carried out – or, in the case of data centre personnel, in close collaboration with the creators of the data – and conduct all or a number of the [data author, data manager, and data user] functions... In origin and training they may be domain experts, computer scientists or information technologists and their career development may have required them to assimilate skills from a discipline from which they did not originate…
- *Data managers:* people who are computer scientists, information technologists or information scientists and who take responsibility for computing facilities, storage, continuing access and preservation of data…
- *Data librarians:* people originating from the library community, trained and specialising in the curation, preservation and archiving of data. Originally, the term data librarian seemed to be confined to librarians dealing with social science data, but the title now encompasses people with data skills in all disciplines… (pg. 8)

As Swan and Brown note, the boundaries between these roles may overlap, with certain individuals taking on more than one role. The responsibility for metadata creation can be equally as fuzzy. Data scientists, managers, or librarians are typically tasked with the job of creating metadata for shared data repositories, but these positions are far from ubiquitous in research settings. In practice, data creators are often expected to create metadata for their data without training or help.

As Edwards, et al. note, "data are the product of 'working epistemologies' that are very often particular to disciplinary, geographic, or institutional locations" [7]. Metadata representations created by data authors are likewise products of 'working epistemologies', and are enacted in different ways in different situations. The standardization of metadata practices varies on a discipline-to-discipline basis. For example, astronomers have made substantial progress in developing community data and metadata standards [10], while habitat ecology has been less successful in this endeavor [2, 13]

Part of the challenge in developing data and metadata systems for research data is that data authors often have little experience in creating structured metadata. Effective information system design can mitigate some of the difficulties resource authors may encounter while creating metadata [4], but research data challenge the metadata creation process in ways that other digital resources, such as web pages and digital documents, do not. The next section introduces our work in designing a metadata repository for data collected by researchers in a large academic science and technology collaboratory.

3. RESEARCH CONTEXT

The research reported here take place within the Center for Embedded Networked Sensing (CENS). CENS is an ideal setting in which to study the emergent changes in scientific research that are being brought about by advanced technology. CENS is a distributed research center [3] based at UCLA with five partnering institutions in central and southern California. Over 200 faculty members, students, and research staff from a number of disciplines are associated with CENS at any given time. The main focus of CENS is to develop sensing systems for real-world scientific and social applications through collaborations between seismologists, terrestrial ecologists, aquatic biologists, and computer scientists and engineers. CENS was founded in 2002 for an initial five years, and received renewal funding in 2007 for an additional five years. Other members of the center come from such disparate disciplines as urban planning, design and media arts, and information studies.

As the center has matured, CENS has been more proactive in making research products available. This has stemmed from internal needs, including the administrative need to keep better track of the center's growth, as well as from external pressure from the NSF to increase the visibility of the center's research output. The first effort in this direction was to make the center's research publications available on the web through the University of California eScholarship repository [15]. This process is still ongoing, but has been largely successful in increasing the visibility and utilization of CENS publications.

The second thrust in our work focuses on research data. Data are taking on growing importance as a product of institutionalized research [1]. We are currently working with the CENS administration to develop a "data sharing" system. The system is being designed to enable re-use and re-purposing of CENS research data by potential outside users. The system will not be collecting the data themselves, due to the marked heterogeneity of data resources and data collection practices of the CENS community [2]. Rather it will focus on making CENS data visible and discoverable to the public by collecting metadata descriptions. CENS researchers will be asked to "register" data using a set of descriptive fields. The data descriptions will then be posted on the CENS website, allowing them to be discovered through web searches or by visitors to the CENS site by provide descriptive information about the data, as well as ways for interested users to get in contact with the appropriate person at CENS for more information.

Our goals in developing a data sharing system for CENS are:

1. *Make CENS data discoverable.* We focus on data "discoverability" because individual/lab data collection and storage practices vary widely within the center. CENS researchers collect a large variety of data resources, including images, audio files, physical samples, and numeric data in both digital and analog form. These resources are spread around many different community, lab, and individual computer systems. Some CENS data is available online through lab websites, but large amounts of data are not currently available online. Because of this variability, collecting and integrating all of the center's data into a single system would be prohibitively expensive and time consuming. Instead, we are designing a metadata repository that allows potential data users to find what data exists and whether the data might be useful to them, as well as providing details about how to get access to data if desired, through links to data or through contact information for the relevant researchers. We are investigating possible policies regarding the timeline for contribution to the metadata system, such as one year from collection or one year from initial funding.
2. *Help CENS researchers keep track of data resources.* In addition to providing a tool that makes CENS data more visible to individuals outside of the center, the metadata repository is intended to help individual research teams within the center to keep track of data created by their own group. Similarly, the metadata repository will provide the center's administrators with a new tool to illustrate the research output of the center.
3. *Sustainability.* The metadata collection system should be sustainable beyond the funding of the center, which will end in 2012. Thus, we are focusing on designing the system to be lightweight, in that it should be easy to use with minimal assistance. Additionally, we are using open source software tools for the back end database and web display.

As an initial step in the design process of this system, we created a preliminary metadata schema that could be used to test possible data description fields. The next section describes a pilot test of these metadata fields, including the fields tested and the test method.

4. TEST METHOD

Four CENS researchers have taken part in the pilot test of the metadata fields, two computer scientists, an engineer, and a domain scientist. The domain scientist and one of the computer scientists are part of the same research team. The participants in this test were chosen through targeted sampling of individuals who were known to have participated in original data collection, and as well as to sample from multiple disciplines and projects within the center. We asked these researchers to create metadata using the below fields for the main data that they were using in their primary day-to-day research. We used a "talk-aloud" protocol, asking the testers to describe what they were thinking and writing as they completed the metadata descriptions. During the test, we observed and took notes of the researchers' activities and comments as they completed the task.

After the testers completed the form, we asked targeted questions about their experience in performing the task. Post-test questions included asking the researchers which fields they felt were the most and least useful in describing their data, what additional fields might be necessary, and what benefits (if any) they feel that they receive from creating this metadata, among others.

The metadata fields used in this test are based on the Dublin Core metadata set [6], as shown in Table 1. The Dublin Core metadata set was chosen for it's flexibility and simplicity in providing descriptive fields for resource discovery. Discipline-specific metadata schemas were considered, such as the Ecological Markup Language and SensorML, but these were deemed to be too inflexible for the diversity of research and data types found in the center.

Table 1. Metadata Fields Used Preliminary Test

Data description fields	Dublin Core elements*
1. CENS project name	title
2. CENS research group	publisher
3. dates (of data collection)	date
4. place	coverage
5. people	-
- contact person	creator
- other participating researchers	contributor
6. data type	type
7. data description	-
- research question (why collected)	description
- what collected (variables)	description
- data collection process and equipment	description
- size, format	format
8. related publications (eScholarship URL)	relation
9. related deployment info. (CENSDC URL)	relation
10. keywords	subject
11. location of the data (URL)	identifier
12. permissions	rights
13. funding source	source

*the Dublin Core element "language" is not used

All of these fields were presented to the user as free-text entries, except for two fields, the "CENS research group" field and the "data type" field, for which the testers were asked to choose from a pre-defined list. The option list for the "CENS research group" field were taken from an established set of categories that exists within the center, and the option list for the "data type" field were taken from the list of data types given in the DCMI type vocabulary. The tests and follow-up questions took between 20 and 30 minutes per tester. The next section describes the main

points of interest that arose during these pilot tests and our subsequent analysis.

5. PRELIMINARY RESULTS

The preliminary findings of the pilot test identify a number of issues that complicate the metadata creation task. Due to the limited scope of this pilot test, these are not meant to be definitive results; rather, they outline important issues that we will use as points for further investigation as our project matures.

- *Item in hand vs. distributed objects:* In many CENS projects data are not individual self-contained items. They may have many constitutive pieces, such as multiple files and database tables, and they may be spread around multiple locations, such as lab servers and personal computers, or for one pilot tester, even located in multiple institutions. In creating metadata, researchers have to decide what is to be described as part of a single project or data set, and where to draw boundaries between data sets.

- *Non-self-describing resources:* Much of the data collected by CENS researchers are not textual, thus researchers must either create textual descriptions from scratch to describe image, audio, or numeric data, or they else adapt existing text from research publications or technical reports to the data description task.

- *Sense making:* Metadata fields may not make sense to a researcher who has not seen them before. In all four pilot tests, the researchers asked for clarification of what they were expected to include in particular fields, requesting examples or further explanation. Some fields, such as "permissions", were problematic to all testers, while other fields, such as "size and format", were only confusing to individual testers.

- *Projected/reverse sense making:* The potential users and uses of research data are often not obvious, even to the researchers who collected them [2]. Researchers must try to project what ambiguous potential future users will need to make sense of their data. In the pilot tests, one strategy people used was to imagine why potential users might be interested using their data. In contrast, one tester took the strategy of thinking about it from the other direction: his own use of outside data. As he said, "If I was going to use somebody else's data, I would want to know…"

- *Talking vs. writing:* In describing their approach to filling out a specific field, particularly fields that they were less sure about, the pilot testers would "talk through" a field until they were surer about what to include. These verbal discussions about what should or should not go in a given field were not always reflected in what was written down. Often a rich verbal discussion resulted in a brief written statement.

- *Individual knowledge vs. group knowledge:* CENS research takes place in group settings. Individual researchers may not know what to include in certain fields, but do know who in the group to ask. For example, a couple of the pilot testers said that they would need to ask their principle investigator about how to fill out the "funding" and "permissions" fields. Another related issue is that different individuals in the same project may have different perspectives on what the boundaries of the data set are, and what descriptive information should be included. For example, the domain scientist and the computer scientists who are part of the same research team emphasized different parts of the same data. As part of the data description, the domain scientist emphasized the physical work involved in installing research equipment in the field and did not provide many technical details. In contrast, the computer scientist emphasized technical features of the data and the way it was collected, and gave no reference to the field work.

- *State of a project:* Different CENS projects are in different states of completion. Metadata has different importance at different stages of a project. One of our pilot testers is involved in a project that has been collecting the same data for over a year and a half, while another pilot tester has been involved in his current project for about six months, with data collection taking place for less than half of that period. In the latter case, the tester described how creating metadata for our system does not benefit him much currently, because his data is so limited in scope that it would not be useful to anyone else. He went on further to say that if they were to expand their data collection considerably, which they hoped to do in the future, then our metadata system would be very useful to him as a means to make his data more accessible to outside users. Additionally, at this early stage of the project, they had not produced any publications or reports on the project, which meant that he did not have any existing text on which to draw in creating metadata descriptions.

Reflecting back on our initial goals – to make CENS data more discoverable, to help research groups keep track of their own data, and to develop a sustainable system – a couple of key challenges require further study. First, many of the issues identified above illustrate the lack of expertise that data authors have in metadata creation. This points to the development of training programs and more explanatory metadata creation systems, including examples and fuller descriptions of the metadata fields. Second, the ambiguity of boundaries around data sets and the fluidity of prospective users and uses of data suggest that training material and activities will need guidelines regarding the focus of the metadata creation process. Third, the tensions between individual and group knowledge suggest that we investigate metadata creation methods that include both individual and group contributions. And fourth, the varied states of project maturity suggest that we investigate the ways that metadata are, or can be, created piece-by-piece during the lifetime of a project.

6. FUTURE DIRECTIONS

Our work on this project is ongoing. We hope to have the initial metadata collection system online by February. At the conference we will present results from this pilot test, as well as results of further tests that take place in the interim period. Over the longer term, we plan to extend this study of metadata creation by data authors in a number of ways. We plan to perform targeted interviews with data creators focusing on understanding their

current metadata practices in their own work. We will ask researchers what "metadata" means to them, what their current data description practices are, and what is involved in sharing their data with people both inside and outside their research group (including the role of metadata in that process). As part of this, we will ask to see the data organization schemes (folder structures, naming conventions, database layouts, etc) currently used by research teams and perform content analysis on these schemes. This will help to characterize the typical state of personal data archives in distributed research environments.

Second, to continue the present study, we are forming plans to use video-taping as a research method. Video-taping will enable more careful study of researchers as they use the metadata creation system introduced in this paper. This will allow us to perform more grounded analysis as our study increases in scale beyond the handful of testers included in our pilot study. Additionally, we plan to investigate new methods of group-oriented metadata creation in our community. Video-taping will be useful in collecting and analyzing the complex interactions that take place in group settings.

These preliminary findings point to the potential contributions of a larger study of the metadata creation process for data creators, including an understanding of:

- the practical difficulties in situations where data managers do not exist for data creators when creating metadata for contribution to a data sharing system
- how metadata creation tasks are parceled out in research groups
- how the setting for metadata creation activities (i.e. individual vs. group) impacts how those activities are conducted

7. CONCLUSION – IMPLICATIONS FOR iSCHOOL RESEARCH

The implications of our research on metadata creation in cyberinfrastructure projects are multi-fold for the iSchool research community, and we hope to promote discussion of these issues. Data are a growing component of the scholarly information infrastructure and must be integrated into larger discussions of technology, institutions, practices, and policy [1]. iSchool research has focused much more on documents than on data. Techniques that have been effective in promoting access and interoperability of documents may not be applicable to data and other digital scientific resources. Research relating to scientific data practices and data preservation and curation are small but growing areas of iSchool expertise. The development of a larger research base in these areas is critical to enhance our understanding of the cyberinfrastructure "blank canvas" [8], and to facilitate the development of a trained workforce of individuals with expertise in data and metadata management [12].

8. ACKNOWLEDGEMENTS

CENS is funded by National Science Foundation Cooperative Agreement #CCR-0120778, Deborah L. Estrin, UCLA, Principal Investigator; Christine L. Borgman is a co-Principal Investigator.

9. REFERENCES

[1] Borgman, C.L. 2007. Scholarship in the Digital Age. Cambridge, MA: MIT Press.

[2] Borgman, C.L., Wallis, J.C., Mayernik, M.S., and Pepe, A. 2007. Drowning in Data: Digital Library Architecture to Support Scientists' Use of Embedded Sensor Networks. In JCDL '07: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (Vancouver, BC). ACM. http://repositories.cdlib.org/cens/wps/216/

[3] Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., & Olson, G. 2007. From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories. Journal of Computer-Mediated Communication. 12(2): 652–672. doi:10.1111/j.1083-6101.2007.00343.x

[4] Crystal, A., and Greenberg, J. 2005. Usability of a metadata creation application for resource authors. Library & Information Science Research, 27(2): 177-189.

[5] Cyberinfrastructure Vision for 21st Century Discovery. 2007. Washington, D.C.: National Science Foundation. http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

[6] Dublin Core Metadata Initiative. 2009. Dublin Core Metadata Element Set, Version 1.1. http://dublincore.org/documents/dces/

[7] Edwards, P.N., Jackson, S.J., Bowker, G.C. and Knobel, C.P. 2007. Understanding Infrastructure: Dynamics, Tensions, and Design. Final report of the workshop, "History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures" [pg. 32]. http://hdl.handle.net/2027.42/49353.

[8] Freeman, P. A., Crawford, D. L., Kim, S., and Munoz, J. L. 2005. Cyberinfrastructure for Science and Engineering: Promises and Challenges. Proceedings of the IEEE, 93(3): 682-691.

[9] Gray, J., Liu, D.T., Nieto-Santisteban, M., Szalay, A., DeWitt, D., and Heber, G. 2005. Scientific Data Management in the Coming Decade. CTWatch Quarterly, 1(1). http://www.ctwatch.org/quarterly/articles/2005/02/scientific-data-management/

[10] Hanisch, R.J. 2006. Data standards for the international virtual observatory. Data Science Journal, 5: 168-173. http://www.jstage.jst.go.jp/article/dsj/5/0/168/_pdf

[11] Lynch, C. 2008. Big data: How do your data grow? Nature, 455(7209): 28-29. http://dx.doi.org/10.1038/455028a

[12] Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century. 2005. Washington, D.C.: National Science Foundation, National Science Board. http://www.nsf.gov/pubs/2005/nsb0540/

[13] Millerand, F. and Bowker, G.C. 2009. Metadata standards: trajectories and enactment in the life of an ontology. in Martha Lampland and Susan Leigh Star (eds) Standards and Their Stories [pp. 149-165], Ithaca, NY: Cornell University Press.

[14] Palmer, C.L., Heidorn, P.B., Wright, D., and Cragin, M.H. 2007. Graduate Curriculum for Biological Information Specialists: A Key to Integration of Scale in Biology. The International Journal of Digital Curation, Volume 2, Issue 2. http://www.ijdc.net/index.php/ijdc/article/viewFile/42/27

[15] Pepe, Alberto, C.L. Borgman, J.C. Wallis, and M.S. Mayernik. 2007. Knitting a fabric of sensor data and literature. in Information Processing in Sensor Networks. 2007. Cambridge, MA: Association for Computing Machinery/IEEE.

[16] Swan, A. and Brown, S. 2008. The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs. Report to the JISC, School of Electronics & Computer Science, University of Southampton. http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf

Extraction and Parsing of Herbarium Specimen Data:
Exploring the Use of the Dublin Core Application Profile Framework

William E. Moen
College of Information
University of North Texas
Denton, TX
940-565-2473
william.moen@unt.edu

Jane Huang
College of Information
University of North Texas
940-565-2473
jqhuang@verizon.net

Melody McCotter
College of Information
University of North Texas
940-565-2473
melodymcotter@gmail.com

Amanda Neill
Botanical Research Institute of Texas
Fort Worth, TX
817- 332-4441
aneill@brit.org

Jason Best
Botanical Research Institute of Texas
Fort Worth, TX
817- 332-4441
jbest@brit.org

ABSTRACT
Herbaria around the world house millions of plant specimens; botanists and other researchers value these resources as ingredients in biodiversity research. Even when the specimen sheets are digitized and made available online, the critical information about the specimen stored on the sheet are not in a usable (i.e., machine-processible) form. This paper describes a current research and development project that is designing and testing high-throughput workflows that combine machine- and human-processes to extract and parse the specimen label data. The primary focus of the paper is the metadata needs for the workflow and the creation of the structured metadata records describing the plant specimen. In the project, we are exploring the use of the new Dublin Core Metadata Initiative framework for application profiles. First articulated as the Singapore Framework for Dublin Core Application Profiles in 2007, the use of this framework is in its infancy. The promises of this framework for maximum interoperability and for documenting the use of metadata for maximum reusability, and for supporting metadata applications that are in conformance with Web architectural principles provide the incentive to explore and add implementation experience regarding this new framework.

General Terms
Standardization

Keywords
Metadata application profiles, Darwin Core, Dublin Core Application Profile, Singapore Framework, biodiversity information, herbarium specimen

1. INTRODUCTION AND RESEARCH PROBLEM
Millions of specimens in museums and herbaria worldwide need to be digitized to be accessible to scientists. Digitizing collections in a well-planned and standard way can increase use and exposure of collections to a more heterogeneous audience while simultaneously reducing physical handling and producing a permanent digital archive [1]. Digitizing the specimen is a necessary but insufficient step to provide effective access and use of the specimen. Converting the specimen metadata into machine-processible form is essential for semantic searching via search engines, distributed databases, and other data portals. A key challenge faced by all natural history collections is determining a transformation process that yields high-quality results in a cost- and time-efficient manner.

The Texas Center for Digital Knowledge in the College of Information at University of North Texas and the Botanical Research Institute of Texas are exploring workflow and metadata issues to design and implement a high-throughput system that exploits computer-assisted human parsing and transformation into structured metadata of herbarium specimen label data. This two-year (December 2008 – November 2010) research projects is funded through a National Leadership Grant awarded by the U.S. Institute of Museum and Library Services. This paper addresses our work to date on metadata issues, and in particular, the use of new frameworks for metadata application profiles.

Herbaria are special natural history collections of preserved plant specimens created for scientific use. Holmgren et al. estimated approximately 3,000 herbaria in 145 countries, containing nearly 300 million specimens [2]. Ongoing collection activities continue to add specimens to existing herbaria. Herbarium specimens are ideal natural history objects, as the plants are pressed flat and dried, and mounted on individual sheets of paper of standard size creating a nearly two-dimensional object. Each specimen is accompanied by a range of information contained on the specimen sheet: attached label with data about the specimens themselves, including the scientific name, where they were collected and by whom and when, and who identified them, as well as other associated data, such as the name of the owning institution or collection, history of ownership, and information added during curation including geocoordinates, as well as measures of data quality [3]. Thus, the specimen sheet contains a wealth of information of interest to researchers, and our project is working to take this unstructured information and transform and enhance it into structured metadata records.

The volume and heterogeneity of the data are challenging for the digitization effort. For example, the Botanical Research Institute of Texas holds over one million plant specimens from around the globe. A survey was made of the complete holdings of one genus, *Artemisia* (sagebrushes and wormwoods), in the Asteraceae (Sunflower plant family). *Artemisia* represented an average holding for the herbarium in terms of size (1179 specimens, or slightly over one cabinet-full), range of localities (worldwide but mostly North America and Europe) and ages of specimens (1805-2007). Only 41% of the *Artemisia* specimen labels were found to be easily machine-readable with off-the-shelf optical character recognition (OCR) software. These specimens were generally North American in origin and collected after 1950. The remaining 59% of specimen labels when processed through OCR resulted in

text containing numerous errors (34%) or were handwritten and impossible to digitize without human processing (25%). Figure 1 presents a sample of the variation in the specimen labels and indicates the challenges to machine-only processes for transformation.

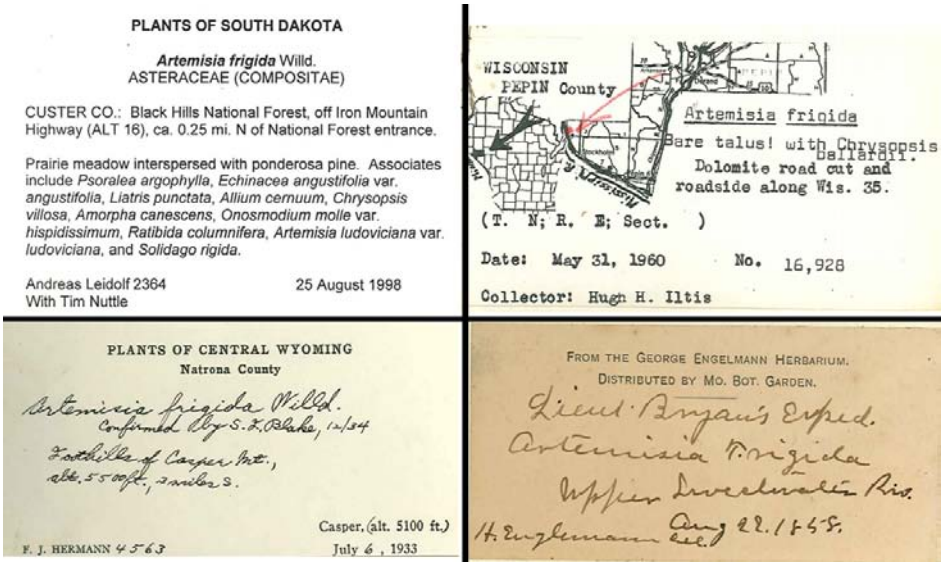


Figure 1. Typical herbarium specimen labels for *Artemisia frigida* from 1998, 1960, 1933, and 1858

2. PROJECT GOAL AND RESEARCH QUESTIONS

The current project has the following overarching goal: **Determine a workflow that provides for a combination of machine-assisted and human-assisted procedures to most effectively and efficiently convert textual data on specimen labels into machine-processible parsed data to ingest in a database and associate with the digitized specimen?** The study is examining how machines and humans can assist each other to yield high-quality and efficiently transformed specimen label data. The central focus of the research is the workflow processes for the transformation of the label data.

Three research questions are addressed in the project: to be addressed are:

- RQ1: To what extent can machine-processes accurately transform label data from a test set of specimen labels that represents variation in label types, quality, and other characteristics (e.g., handwritten versus typescript)?
- RQ2: Which human processes can be incorporated into a robust workflow to further transform, correct, and enhance label data?
- RQ3: What user interfaces are most effective and suitable to the tasks and users in supporting human processes in the workflow?

The results of this research will yield a new workflow model for effective and efficient label data transformation, correction, and enhancement that can be replicated, adapted, and transferred to herbaria and other natural history collections. Project activities are underway to address these research questions, and reports and papers generated later in the project will provide our answers to these questions. The remainder of this paper discusses metadata aspects of this project.

Metadata plays several important roles in this project. Of primary importance is capturing all relevant information from the specimen sheet, structuring that information into a specimen record in a format that uses appropriate terms from existing metadata vocabularies to ensure the metadata is shareable and enables interoperability and integration with other systems and applications. Metadata is also being used to support workflow processes and manage the digitized specimen image and derivative objects as they move through the workflow.

The following objectives related to metadata are being addressed in this project:

- Determining the metadata requirements to support the workflow and the specimen label data
- Identifying appropriate metadata vocabularies to use
- Formalizing and documenting the metadata used in the project to increase the shareability and interoperability of the resulting metadata records

3. METADATA FOR BIODIVERSITY INFORMATION

The communities working with biodiversity information, which includes botanists and herbaria, have been evolving metadata and other standards over the 8-10 years. Early work on metadata for biodiversity was initiated as part of the Species Analyst Project (http://xml.coverpages.org/speciesAnalyst.html). Emerging from

that work was a metadata scheme called Darwin Core. Since that time, the Taxonomic Database Working Group (TDWG, now referred to as the Biodiversity Information Standards group) has evolved the Darwin Core (DwC), and in October 2009 ratified a new version of DwC as a TDWG standard; DwC was developed to facilitate the sharing of information about biological diversity. (http://rs.tdwg.org/dwc/terms/index.htm). DwC terms focus on taxa, their occurrence in nature as documented by observations, specimens, and samples, and related information (http://rs.tdwg.org/dwc/index.htm).

The Dublin Core Metadata Initiative (DCMI) served as a model for DwC, and DwC can be viewed as a general extension to Dublin Core (DC) metadata terms. DwC uses a number of DC terms and also defines a list of terms that address the information needs of the biodiversity information community. With the ratification of the DwC standard, the community now has a solid basis on which to develop metadata records describing a broad range of naturally occurring organisms whether at the macro or micro level (e.g., animals to genes).

Since our current research project deals with herbarium specimens and associated data, the project is using DwC as the foundational metadata vocabulary. Specifying the use of DwC metadata as well as accommodating the needs of the project for metadata beyond what DwC offers requires a method for using and documenting metadata terms from various namespaces (i.e., from other metadata vocabularies).

4. METADATA APPLICATON PROFILES

The concept of application profiles has evolved in the past 10 years. Heery and Patel [4] first proposed profiles as a method for documenting the use, in a single application, of metadata elements from various namespaces. Application profiles can specify the use of, and constraints on, metadata elements in particular applications. In 2003, a European Committee for Standardization Workshop resulted in the *Dublin Core Application Profile Guidelines* (ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14855-00-2003-Nov.pdf). The form of these application profiles were typically documents that could be used by both producing and consuming applications. On the producing side, the application profile guided the input requirements for the creation of metadata records. For example, the application profile indicated the elements that would be in the metadata record, obligations and constraints on individual elements (e.g., whether an element was mandatory, repeatable, and/or used data values from specific controlled vocabularies). For those consuming or using the metadata records, the application profile provided the details to a system developer to know what to expect in the metadata record and thus develop programs to ingest and make sense of the metadata. The limitation of this approach to application profiles was that the profile document was not machine-actionable. It typically took the form of a text document.

More recently, the Dublin Core Metadata Initiative (DCMI) proposed a Dublin Core Application Profile (DCAP) framework "for maximum interoperability and for documenting such applications for maximum reusability". DCAPs developed using this new framework are intended to support metadata applications that are in "conformance with Web-architectural principles," and in particular, serve the needs of the Semantic Web (http://dublincore.org/documents/singapore-framework/).

The following sections describe the work to date in our project to develop and implement an application profile using this new framework.

5. THE NEW DUBLIN CORE APPLICATION PROFILE FRAMEWORK

Just as DwC metadata has evolved over the years to meet the needs of the biodiversity information community, the DCMI has also evolved along several dimensions: terminology, concepts, models, and support for emerging semantic web technologies. A key moment in this evolution was the adoption in 2005 of the Dublin Core Abstract Model (DCAM) with a status "Recommended". The abstract model was intended to "specify the components and constructs used in Dublin Core metadata… [and define] the nature of the components used and describes how those components are combined to create information structures (http://dublincore.org/documents/abstract-model/). The resulting information model was not tied to a particular encoding syntax, and instead was intended to assist understanding of the kinds of descriptions being created.

The DCAM defines three related model: Resource Model, Description Set Model, and Vocabulary Model. For example, the Resource Model is represented in Figure 2 with text explanation following.

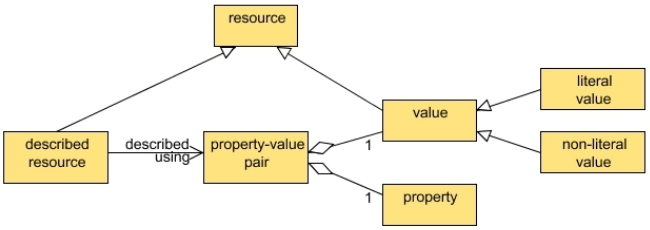


Figure 2. DCAM Resource Model

The abstract model of the resources *described by* descriptions *is as follows:*

- *Each* described resource *is described using one or more* property-value pairs.
- *Each* property-value pair *is made up of one* property *and one* value.
- *Each* value *is a* resource - *the physical, digital or conceptual entity or* literal *that is associated with a* property *when a* property-value pair *is used to describe a* resource. *Therefore, each* value *is either a* literal value *or a* non-literal value*:*
  - *A* literal value *is a* value *which is a* literal.
  - *A* non-literal value *is a* value *which is a physical, digital or conceptual entity.*
  - *A* literal *is an entity which uses a Unicode string as a lexical form, together with an optional language*

tag or datatype, to denote a resource *(i.e. "literal" as defined by RDF).*
(http://dublincore.org/documents/abstract-model/).

At the 2007 *International Conference on Dublin Core and Metadata Applications* in Singapore a new framework for application profiles was proposed. The framework, *Singapore Framework for Dublin Core Application Profiles* (http://dublincore.org/documents/singapore-framework/) builds upon the concepts described in the DCAM, and defines a formal mechanism for creating DCAPs "for maximum interoperability and for documenting such applications for maximum reusability". The framework defines the necessary components of an application profile:

- ***Functional Requirements (mandatory)*:** Describes the functions that the application profile (AP) is designed to support, as well as functions that are out of scope.
- ***Domain Model (mandatory)*:** Describes the objects metadata will describe and the relationships between those objects.
- ***Description Set Profile (DSP) (mandatory)*:** Defines a set of metadata records that are valid instances of an AP.
- ***Usage Guidelines (optional)*:** Describes how to apply the AP, how the properties are intended to be used in the application context, etc.
- ***Encoding syntax guidelines (optional)*:** Describes AP-specific syntaxes and/or syntax guideline, if any.

Figure 3 indicates the relationship of the components of the application profile with other related resources.

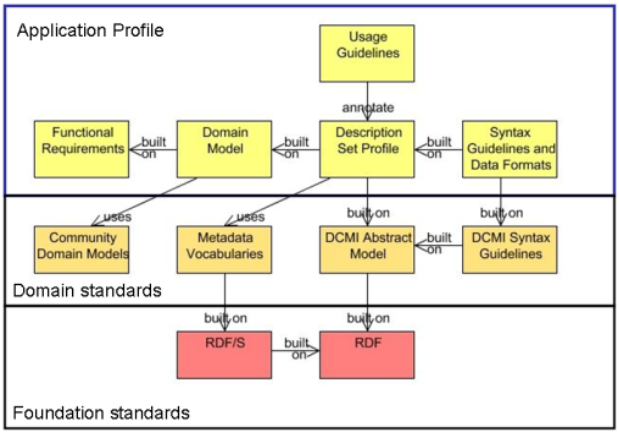


Figure 3. Application Profile Model

(http://dublincore.org/documents/singapore-framework/)

The development of application profiles in the new DCAP framework can be considered to be in its infancy with few examples from which to draw (e.g., the Dryad Project, see Greenberg, et al., 2009 [5]; Scholarly Works Application Profile, http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile).

6. THE PROJECT'S APPLICATION PROFILE RENDEREDIN THE DACP FRAMEWORK

Exercising the DCAP framework in the context our project provides a way of better understanding just what this framework requires as well as the benefits that can accrue from this approach. The following discussion is based on our work so far and may provide clarification to others who are considering using the DCAP framework. In developing the DCAP we are relying on three key documents:

- *The Singapore Framework for Dublin Core Application Profiles*: http://dublincore.org/documents/singapore-framework/
- *Guidelines for Dublin Core Application Profiles*: http://dublincore.org/documents/profile-guidelines/
- *Criteria for the Review of Application Profiles*: http://dublincore.org/documents/profile-review-criteria/

The Singapore Framework for Dublin Core Application Profiles (described above) identified several component parts of a DCAP; the *Criteria for the Review of Application Profiles* document adds one more component that addresses the objectives and scope of the application. The following sections address the three mandatory components of a DCAP.

6.1 DCAP Functional Requirements

The preliminary workflow for extracting and parsing of specimen label data is represented in Figure 4.

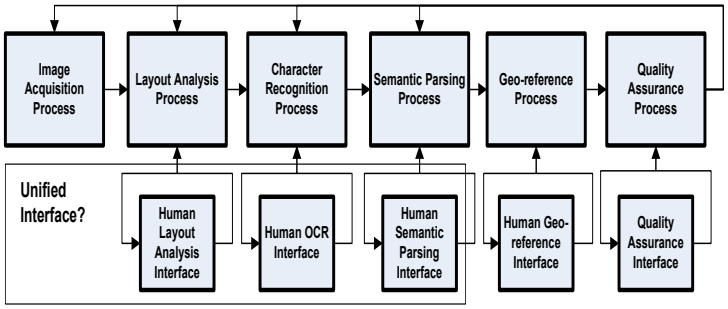


Figure 4. Transformative Process Workflow

This provides a starting point for thinking about the metadata needs that the application profile will address.

The project's functional requirements include high-level system requirements and goals (e.g., optimizing the workflow, system integration, and reusability of code) as well as more detailed requirements, especially in terms of metadata needed for various objects that move through the workflow in Figure 4. We used processed-centered use case modeling to identify key objects, subprocesses, and tasks for each of the processes outlined in

Figure 4. Specific metadata requirements that have been identified relate to: types of metadata; standard vocabularies (e.g., DwC); consistency and comprehensiveness; interoperability/shareability; granularity; reusability; and specific constraints on metadata terms.

6.2 DCAP Domain Model

This aspect of the DCAP relates to the objects of interest to the application profile and the metadata associated with each. The domain model for the project includes and defines four objects within the workflow that require metadata and shows the relationships/derivations of the separate objects. The four objects are:

- **Specimen Object:** This will have metadata derived from all of the information from the specimen sheet.
- **Specimen Image Object:** A scan of the herbarium specimen sheet and the source from which ROIs are derived.
- **Region of Interest Object (ROI):** A ROI is derived from the specimen image object and can include separate ROIs for primary label, first annotation, and other textual or graphical information on the herbarium sheet.
- **Digital Text Object:** This object results from OCR processing of a ROI or manual transcription of data from an ROI.

Relationships between these objects can be one-to-one (1…1) or one-to-many (1…n). Figure 5 shows the four objects and the relationships between the objects.

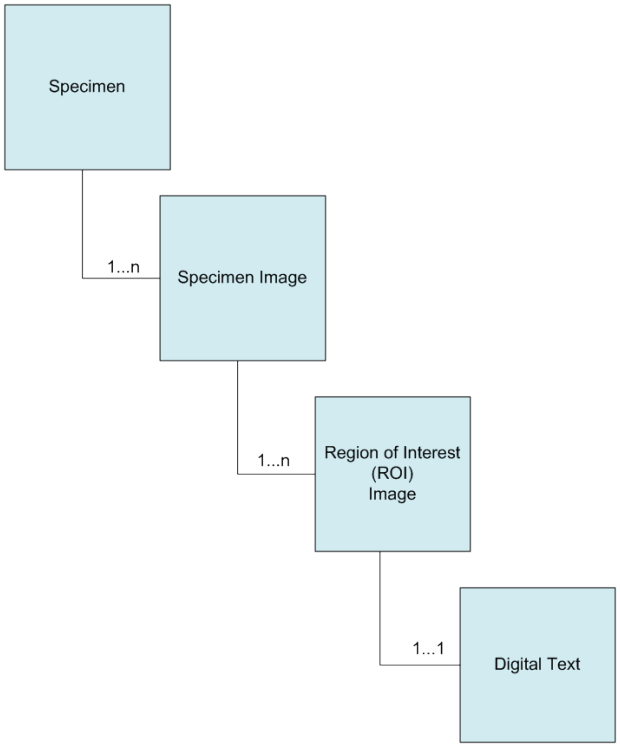


Figure 5. Objects in the Project's Domain Model

6.3 DCAP Description Set Profile

The description set profile (DSP) serves a key function in the application profile for defining the metadata terms that will be used and constraints on the use of the terms. Figure 2 above shows that the DSP is built upon domain standards that include metadata vocabularies. We see the development of the DSP is at least a two-step process:

- Determining the metadata terms required
- Formalizing the use of the terms in a structured document.

6.3.1 Determining the metadata terms

Darwin Core provides a community and domain standard for metadata terms that will be used in the application. However, from an analysis of the information on the herbarium specimen sheets we are using in the project (approximately 1,000 type specimens), DwC does not appear to accommodate all the information that appears on the sheets and that need to be recorded in the specimen metadata record. In the early phase of the project (Spring 2009), we identified a set of elements needed to accommodate the information needs. We then did a mapping to the existing DwC terms. Since the DwC was approved in October 2009 as a ratified TDWG standard, we are again investigating the DwC terms that can be used for the project's needs. For those that are not available in DwC, we need to locally define in a new namespace the terms needed.

Although the specimen label data are the focus of the specimen metadata record, the workflow also requires some technical and other metadata to help manage the objects as they move through the workflow. Two likely sources of terms are the Metadata for Images in XML (http://www.loc.gov/mix/) and Preservation Metadata (http://www.loc.gov/standards/premis) vocabularies.

6.3.2 Formalizing the use of the terms in a structured document

The concept of a description set profile model was first articulated in the DCAM document. In 2008, the DCMI published a more complete articulation of the concept in *Description Set Profiles: A Constraint Language for Dublin Core Application Profiles* (http://dublincore.org/documents/dc-dsp/). According to this document, a "DSP is a way of describing structural constraints on a description set. It constrains the resources that may be described by descriptions in the description set, the properties that may be used, and the ways a value surrogate may be given."

In the tradition "mix and match" approach for application profiles described by Heery and Patel [4] the metadata terms used in an application and constraints could be represented as in Table 1. We will use two metadata terms from our project to illustrate.

Term URI	http://rs.brit.org/ap/terms/barcode
Defined by	http://rs.brit.org/ap/
Name	Barcode
Source definition	The verbatim supplemental text associated with a barcode imprinted or affixed to the

	specimen.
Local definition	The verbatim supplemental text associated with a barcode imprinted or affixed to the specimen.
Type of term	n/a
Refines	n/a
Has encoding scheme	No
Obligation	Optional
Occurrence	Non-repeatable
Datatype	String

Term URI	http://rs.tdwg.org/dwc/terms/#scientificName
Defined by	http://rs.tdwg.org/dwc/terms/
Name	scientificName
Source definition	The taxon name (with date and authorship information if applicable). When forming part of Identification, this should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the IdentificationQualifier term.
Local definition	The taxon name (with date and authorship information if applicable). When forming part of Identification, this should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the IdentificationQualifier term.
Type of term	n/a
Refines	Has domain: http://rs.tdwg.org/dwc/terms/#Taxon
Has encoding scheme	http://www.ipni.org/ or BRIT compilation
Obligation	optional
Occurrence	repeatable
Datatype	string

Table 1. Recording Information for Metadata Terms in Traditional Format

Using the new DCAP framework, the above specifications can be rendered in a Description Set Profile. For ease of reading, we present the terms and constraints in a human-readable format as follows:

DescriptionSet: SpecimenData

Description template: Specimen
minimum = 1; maximum = 1
Statement template: barcode
minimum = 1; maximum = 1
Property: http://rs.brit.org/ap/terms/barcode
Type of Value = "literal"
Statement template: scientificName
minimum = 0; maximum = unlimited
Property: http://rs.tdwg.org/dwc/terms/#scientificName
Type of Value = "non-literal"
Take list = yes
Value Encoding Scheme URI = http://www.ipni.org/

The above indicates that the Description Set is related to something called a specimen (i.e., the botanical specimen on the herbarium sheet). The Description Template provides information about each object in the Domain Model. The statement "minimum = 1; maximum = 1" means that the metadata record represents one and only one specimen. The Statement Template contains the statements about the properties (metadata terms) used to represent the Specimen, giving information about the number of occurrences a term can have in the record, the URI to the property (metadata term), type of value associated with the term, and other constraint information.

The DSP can be represented in XML as well as RDF. The following shows the above information represented in XML

```
<DescriptionSetTemplate>
<DescriptionTemplate ID="Specimen" maxOccur="1" minOccur="1" standalone="no">
<ResourceClass>http://rs.brit.org/ap/objects/SpecimenMetadata</ResourceClass>
<StatementTemplate ID=”barcode” minimum= “1” maximum= “1” type="literal">
<Property>http://rs.brit.org/ap/terms/barcode</Property>
<LiteralConstraint>
<SyntaxEncodingSchemeOccurance>disallowed</SyntaxEncodingSchemeOccurance>
<LanguageOccurance>optional</LanguageOccurance>
</LiteralConstraint>
</StatementTemplate>
<StatementTemplate ID=”scientificName” minimum= “0” maximum= “unlimited” type="nonliteral">
<Property>http://rs.tdwg.org/dwc/terms/#scientificName</Property>
<NonliteralConstraint>
<VocabularyEncodingSchemeOccurrence>optional
</VocabularyEncodingSchemeOccurrence>
<VocabularyEncodingSchemeURI>http://www.ipni.org/
</VocabularyEncodingSchemeURI>
<ValueStringConstraint maxOccurs="0"/>
</NonliteralConstraint>
</StatementTemplate>
</DescriptionTemplate>
</DescriptionSetTemplate>
```

A complete DSP will address each metadata term used in the application's metadata record, indicating information about the term (e.g., URI to the definition of the term), and indicating specific constraints on using the term in this particular

application. The DSP representation in XML or RDF provides what was not possible in the earlier forms of application profiles, namely having a machine-actionable representation of the metadata used in a particular application.

7. SUMMARY AND CONCLUSION

The research and development being addressed in our project focuses on two major areas:

- Development and testing of optimal workflows to extract, parse, and enhance the data from herbarium specimen sheets
- Using a standards-based approach for the resulting metadata describing the specimen on the herbarium specimen sheet.

This paper has described how we are exploiting new developments, concepts, and formalisms of the Dublin Core Metadata Initiative to improve the shareability and interoperability of the metadata created through the workflow. Using the Dublin Core Application Profile framework enables machine-actionable application profiles. This should lead to more efficient and effective data integration among systems and applications. The project is connecting concepts and practices of standards-based metadata, shareability and interoperability with the needs and goals of a large herbarium to make the valuable specimen label data available to botanists and other researchers. Providing legacy specimen data in the form of structured botanical metadata records, along with high resolution images of plant specimens, will provide new research opportunities for the biodiversity information community.

8. REFERENCES

[1] National Science Board. 2005. Long-lived Digital Data Collections: Enabling research and education in the 21st century. NSF. (http://www.nsf.gov/pubs/2005/nsb0540).

[2] Holmgren, P.K., N.H. Holmgren and L.C. Barnett. 1990. Index herbariorum. Part I: The herbaria of the world. 8th edition. New York Botanical Garden. 693 pp.

[3] Morris. P.J., 2005. Relational database design and implementation for Biodiversity Informatics. *Phyloinformatics* 7:1-63. (http://www.athro.com/general/Phyloinformatics_7_85x11.pdf).

[4] Heery, Rachel, and Manjula Patel. (2000). Application profiles: Mixing and matching metadata schemas. *Ariadne, 25*. Retrieved November 15, 2009, from http://www.ariadne.ac.uk/issue25/app-profiles/.

[5] Greenberg, Jane, Hollie C. White, Sara Carrier, and Ryan Scherle. (2009). A metadata best practice for a scientific data repository. *Journal of Library Metadata*, 9:3, 194-212.

The Role of Informatics in Software Engineering: Literature Reviews, Agenda and Software Informatics

Ira Monarch
Software Engineering Institute
4500 Fifth Avenue
Pittsburgh, PA 15213
001-412-268-7070
iam@sei.cmu.edu

Sheila Rosenthal
Software Engineering Institute
4500 Fifth Avenue
Pittsburgh, PA 15213
001-412-268-7846
slr@sei.cmu.edu

Rachel Callison
Software Engineering Institute
4500 Fifth Avenue
Pittsburgh, PA 15213
001-412-268-7725
callison@sei.cmu.edu

ABSTRACT
Literature reviews have been and are increasingly being merged with semi-automatic versions of bibliometrics and text analytics to extend library capabilities in the direction of performing informatics. Such extended capabilities allow special libraries to move in the direction of supporting and even performing informatic functions for bioinformatics or medical informatics. Recently a new subfield of informatics, software informatics, has been discussed that opens informatic opportunities for software engineering special libraries. The paper discusses a software engineering library that is using informatic techniques to support characterizations of customer demand landscapes that inform software engineering agenda. Sources analyzed go beyond published periodical literature to include organizational reports like budget justifications and, potentially, use of web harvesting. It is proposed that this use of informatic techniques is part of software informatics and because of the potential impact on how software is developed and used, may be part of software engineering as well.

Categories and Subject Descriptors
D. Software; H.3.1 Content Analysis and Indexing; K.6.3 Software Management.

General Terms
Management, Measurement, Documentation, Design, Human Factors.

Keywords
Software engineering, software informatics, bibliometrics, text analysis, informatics, special libraries.

1. INTRODUCTION
There are many characterizations of informatics. The paper characterizes informatics as the study, use and communication of information including its analysis and organization. Sub-domains of informatics focus on specific domains like chemical-, bio-, medical- geo-, social- business-, library- and recently software-informatics. The sub-domains all employ information technology to manage, process and analyze data and information pertinent to a given domain. Informatic sub-areas also approach information from various perspectives, individual, social, economic and cultural. Employing information technology from these perspectives can help to provide support for constructing and carrying out agendas of various disciplines (e.g., chemical engineering, bio-medicine, business re-engineering and software engineering) in a given domain. Nevertheless, work in informatic sub-domains is not always thought to be part of these engineering, medical and business disciplines, even if the work is thought to be part of a discipline. Moreover, reviews of textual sources and their analyses, and library informatics more generally, are not always thought to be part of the informatics of sub-domains, and even less likely to be thought of as part of the disciplines corresponding to an informatic sub-domain. To open the possibility for extending software informatics, the paper proposes, as a subject for discussion, the idea of using informatic techniques to support learning software engineering agenda and informing those who could influence these agenda.. The paper further proposes, again as a subject for discussion, that this extended version of software informatics be considered part of software engineering.

In order to discuss these proposals, it is important to get clear on the notion of agenda. According to Michael S. Mahoney, an historian of mathematics who also wrote extensively on the history of software engineering and computation, an agenda lies at the heart of a discipline [5]. Here Mahoney is referring to disciplines such as applied mathematics, computer science or software engineering. An agenda for Mahoney is

a shared sense among its practitioners of what is to be done: the questions to be answered, the problems to be solved, the priorities among them, the difficulties they pose, and where the answers will lead. When one practitioner asks another, "What are you working on?" it is their shared agenda that gives professional meaning to both the question and the response. Learning a discipline means learning its agenda and how to address it, and one acquires standing by solving problems of recognized importance and ultimately by adding fruitful questions and problems to the agenda. Indeed, the most significant solutions are precisely those that expand the agenda by posing new questions … Most important for present purposes, the agendas of different disciplines may intersect on what come to be recognized as common problems, viewed at first from different perspectives and addressed by different methods … [5]

This notion of agenda provides a backdrop for much of the discussion in the paper. However, it has typically been the case that only practitioners of a discipline, those on the inside, get to determine the discipline's agenda. Customers or users of what the discipline produces according to its agenda are almost never considered. This would seem to make most sense in pure mathematics whose products are papers containing mathematical proofs, read for the most part, by other pure mathematicians. However, this is not always the case. Even pure mathematics often gets applied and sometimes different disciplines mutually influence each others' agendas as has been the case for abstract algebra and computer science [5 and 7]. Nevertheless, on Mahoney's view, disciplinary agenda are influenced from the

outside only when the outside is another discipline part of whose agenda can be borrowed by the receiving discipline. What is not considered is an outside that is articulated in terms of what a user needs or demands from the discipline. A classic articulation of this view for software engineering is from Dijkstra who believed that software engineering is not predicated on user demands but rather on mathematics [3]. This paper will consider a very different view discussing how laypersons can influence specialist agendas [8] through their document representatives [9].

In what follows we describe a specific case in which informatic-based text mining and analysis is performed with the goal of agenda learning/influencing based on characterizing parts of the customer demand landscape of software engineering. Interspersed in the description of this case, we further discuss the relationships between informatics, software informatics and software engineering.

The next section introduces the Software Engineering Library at the Software Engineering Institute (SEI) as a special library supporting Software Engineering and goes on to discuss its role in document searches and analyses for better understanding of the software engineering demand landscape. This is followed by a discussion of informatic techniques that can be, and to some extent are being, merged with special library functions. These techniques in the form of text analytics applied to the documents collected are used as a basis for understanding the demand landscape and informing the subsequent building of a software engineering roadmap. The final section summarizes the case for the new field of software informatics and its role in software engineering and makes a case for including special library functions, bibliometrics and text analytics in software informatics because of their role in building roadmaps and understanding/influencing agenda.

2. A Special Library Supporting Software Engineering

The definition of a Special library, according to the Online Dictionary for Library and Information Science, by Joan M. Reitz is "a library that is established and funded by a commercial firm, private association, government agency, nonprofit organization, or special interest group to meet the information needs of its employees, members, or staff in accordance with the organization's mission and goals." [6] The SEI Library fits the definition. Established and funded by the SEI in 1986, its mission is to efficiently provide timely and relevant information to the SEI by maintaining an expert library staff and a strong collection in software engineering, computer science and related disciplines.

The SEI Library is staffed by a library manager, a reference librarian, an archivist, and a paraprofessional. In addition, the library currently has the good fortune to have the assistance of a recent Carnegie Mellon art design graduate who has created the information visualization poster shown in Figure 1. This poster is prominently displayed on the wall behind the reference desk in the library, showcasing both the library and the archive. The SEI Archive was created in 2005 under the supervision of the library's departmental Director. The archivist works closely with the librarians and many of the services overlap, as illustrated above. One of the major goals of the SEI Library is to increase its involvement in the research required by the Institute's software engineers. Therefore, the library focuses on the following areas of commitment: encourage collaboration on SEI projects; develop beneficial programs for SEI staff; and provide special services that will enhance library support for their research.

In order to understand the Software Library's mission, the SEI's mission must also be stated. According to SEI's website (http://www.sei.cmu.edu/about/) its mission is to advance software engineering and related disciplines, though with an eye to ensuring that the development and operation of systems is predictable with improved cost, schedule, and quality. This means, according to the terminology of this paper, that the SEI's agenda for software engineering focuses on certain kinds of systems and certain of their attributes. The SEI Library by supporting SEI's mission supports the discipline of software engineering in this sense. In the case to be described the SEI Library was able to participate in performing extended informatic functions that informed SEI agenda learning/influencing. What will be described in the rest of this section and in following sections are the role the library played in performing the informatic functions, whet these functions were and what findings were produced. Any use the ED project or the SEI made of the findings will not be described.

Figure 1: Work of the Software Engineering Institute Library

The SEI Library has participated in a project created by the new Executive Director of Interagency and Cyber Initiatives and head of the Acquisition Support Program (ED-ICI & H-ASP or ED for short) when she joined the SEI Staff in May 2009. Not long after the ED arrived, she informed the library staff that she was interested in everything *cyber* and was starting a new project[1] to locate all current information on this topic. The SEI Library participated in the project through one of its main functions, providing *current awareness*, including *table of contents* and *keyword alert* services.

The journals the ED selected for her table of contents current awareness included: *Defense News Early Bird and Daily News*

[1] This will be called the ED Project in the rest of the paper.

iConference 2010, February 3-6, 2010, University of Illinois at Urbana-Champaign, IL, USA.

Roundup, *The Economist*, and the *Wall Street Journal*. In addition to these publications, the library staff provided her Tech America Headlines from Infoition, which are scanned every morning for information on *cyber*. This news service includes such publications as the *Washington Post, Information Week, NextGov, Wall Street Journal, Federal Computer Week, Business Week, Associated Press, New York Times, Technology Review, Computer World, CNET News, Government Computer News,* and *Time*, providing very thorough daily coverage on this topic.

Keyword alerts for the ED are run in *Academic One File*, which is a service provided by Gale Publications; Elsevier's *E.I. Engineering Village,* which contains two files, *Compendex and Inspec;* and UMI's (University Microfilm International's) *ProQuest Database*. The retrieval from these databases comes to the ED's email address directly, with many providing full text options. The ED shares important key and high impact articles with other members of her new project who include representatives from all SEI Programs, Services, and Functions. As certain specific phrases using the term *cyber* gathered more impact or focus through the media, e.g., the phrase, *cyber command*, the ED would request targeted searches to run on these specific phrases.

After being in her new position for only a few months, The ED created the concept for another type of library alert service. She asked if the library could, on a regular basis, search certain websites for the following specific terms: cyber warfare, cyber assurance, cyber component commander, 24th Air Force, and AF cyber security. In addition and also on a regular basis, the ED requested a similar set of alerts from the OSD Homepages for OSD NII, OSD USD-Intelligence, Office of the Director of National Intelligence, IARPA and DARPA to start.

One of the teams working on the ED Project devised specific literature search strategies for retrieving *cyber* related information focusing in on DOD, civil agencies and the business sector. The team wanted to identify and quantify future business opportunities for software reliant systems in *cyber environments*; establish the dollars to be spent; forecast business risk over the next three to five years for software reliant systems in *cyber environments*; and identify budgetary focus on such items as training, software, infrastructure, systems and platforms. They identified specific budget topics or keywords to query in the context of software reliant systems, software intelligent systems, and systems of systems in *cyber environments*. This team's dual role of identifying and quantifying business opportunities builds a bridge, in effect, between the special library function to support literature searches and the extended library function of using bibliometrics and text analytics to help the ED project interpret the results of the literature searches and make forecasts. The text analysis described in the section after next is applied to some of the documents found in the searches described above.

3. *Bibliometrics* and Text Analytics

A characterization of informatics was provided in the introductory section. Bibliometrics and text analytics will be introduced in this section.

Bibliometrics is a set of methods, usually statistical or mathematical, used to study or measure the attributes and relations of collections of bibliographic references or documents. Bibliometric methods are most often used in the field of library and information science. In this sense bibliometric methods can be seen as part of informatics. Citation analysis and content analysis are commonly used bibliometric methods, though historically bibliometrics has devoted more attention to citation analysis. Bibliometric methods can be used to explore the impact of research fields on one another, the impact of a set of researchers or the impact of a particular paper. In each case, impact on practice is determined by bibliometric analysis of documents made available in other ways than through publishing vehicles such as journals.

Citation analysis has been used to evaluate the importance of a research article, the researchers themselves through their published work or a whole department in a university. Citation analysis is often used to identify the watershed publications in particular disciplines through quantitative means and the interrelationships between authors from different institutions and schools of thought. Such citation analysis depends on citation indices, such as Institute for Scientific Information's (now Thomson Reuters') Web of Science.

Some limitations on the value of citation data have been pointed out, especially in regard to quality ratings. For example, a low correlation has been found between peer evaluation of computer science groups and citation impact indicators of their papers [2]. In addition, the h-index (A scholar has an index of h if he or she has published h papers each of which has been cited by others) can be misleading [2]. These points are reinforced by a report that strongly cautions against over-reliance on citation statistics such as the impact factor and h-index [4].

Content analysis using semi-automated text analytic techniques can avoid some of the issues with citation analysis. It can identify the role that authors play in originating and/or contributing to the thematic structure of a discipline. The emphasis can therefore be on who introduces new ideas and applies them, rather than on sheer citation counts. Moreover, by focusing on the thematic structure of a discipline or an ultra large collection of documents, content analysis, or text analytics, can address questions concerning the *content* of the discipline or document collection that citation analysis cannot. This paper concentrates on the role of text analytics and its application to software engineering literature reviews, road maps and to other parts of software informatics because its value is underappreciated. However, citation analysis used critically either alone or in conjunction with content analysis also has value.

4. Applying Text Analytics to uncover a Software Engineering Demand Landscape

This section describes the application of semi-automated text analysis to two sets of documents being collected in a literature search, partly described above, to gain insights into the current software engineering demand landscape. One set of documents focuses on budget and financial summaries from the DOD, DARPA, DHS, DOE and US Military Services etc. that provide justifications for their technology and service acquisitions; the other focuses on descriptions of technologies and services that are seen as attractive or promising to actual or potential SEI customers. So far, 10,000 pages of budget and financial summaries and over 1000 pages describing promising technologies have been analyzed. The aim of these and other analyses is to support (1) understanding of the demand landscape of potential and actual SEI customers, (2) determining the extent to which current SEI services align with the demand landscape

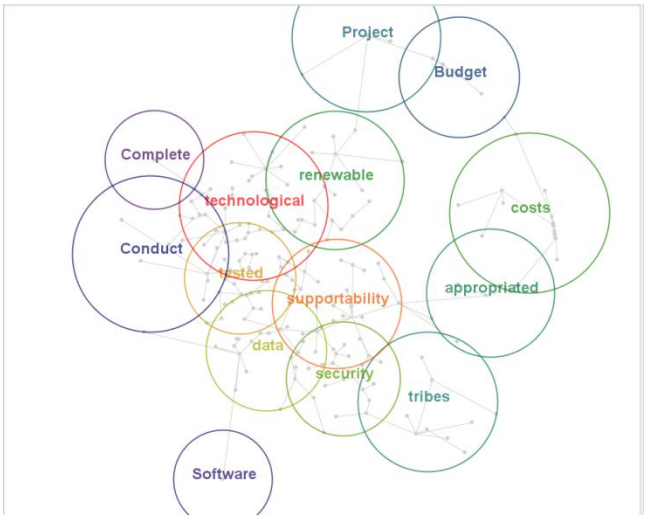
and (3) making suggestions for better SEI alignment with demand and vice versa. This paper will focus mostly on 1 and a little on 3.

Two main questions guided the text analysis. (1) How well are software related concepts integrated into budget justifications of promising technologies that support cyber environments? (2) Do cyber concepts play a mediating role between software related concepts on the one hand and concepts involved in budget justifications and promising technologies on the other? In both cases the short answer, based on the text analysis and interpretation of concept maps, turned out to be only in a limited way. Further, in answer to the second question, the text analytic findings also indicated that the concepts of **data** and **networks** already do play a more important mediating role than **cyber** concepts and in the future could play an even larger mediating role. Approaching engineering agenda in terms of mapping the conceptual space of what is engineered is important because design is a primary activity in engineering, and the design of software, like the design of any artifact, is based on the meaning of an artifact negotiated by all design stakeholders including users and customers as well as engineers [10]. Concept maps generated by automated text analysis techniques can be useful in this regard (given that the right text sources are selected for analysis) to provide a snapshot of the design or "drawing together" of an artifact in all its interrelationships and complexity [10].

Automated text analysis produces what are called concept maps.[2] A concept map generated for the budget and financial summaries is shown in Figure 2. The themes are the large circles with labels inside the circles capturing the theme. The circles representing themes also contain light grey dots or nodes representing concepts (not labeled in Figure 2) and lines representing links between the concepts. Links between concepts are determined on the basis of co-occurrence of terms.[3] Because concepts with different meanings can frequently co-occur, concepts in different themes are often linked. The intensity or thickness of the lines representing the links indicates likelihood of co-occurrence. The themes with the most concepts and links are ranked more highly in terms of Connectivity and overall Relevance to the contents of the collection of documents being analyzed. The top theme in terms of Connectivity and Relevance is ***technological***. It is set at a benchmark of 100% that other themes are measured against. A table showing the ranking of the themes is included under the concept map in Figure 2.

The nodes representing concepts are also labeled (see the concept **technological** in the upper left part of the circle containing the theme ***technological***[4] in Figure 3). Concepts stand for more than the literal words or phrases that label a node. They also stand for an affinity list of terms that strongly co-occur with the term standing for the concept. The term standing for the concept is automatically chosen on the basis of its more strongly co-occurring with all the other terms on the affinity list than any other term on the list. Themes are clusters of concepts that have more similar co-occurrence patterns with each other than concepts in other thematic clusters have with each other. Themes are automatically named by the concept that has a co-occurrence pattern more similar to the co-occurrence patterns of all the other concepts in the cluster than does any other concept in that cluster.

The meaning of a theme derived from a collection of documents, for example ***technological*** in the budget and financial document collection, is determined by the concepts in the cluster that constitute the theme. In the case of ***technological***, this list of concepts includes: **renewable** (technologies, e.g., **ethanol**), **engine, aircraft, hyperspectral/polarimetric** (remote sensing) **vehicles, amplifiers, weapon, biology, catalytic, chemicals, nuclear, medical, environmental, missile, algorithms.**



Thematic Summary

Theme	Connectivity	Relevance
technological	100%	
supportability	92%	
tested	69%	
data	46%	
security	36%	
costs	19%	
renewable	15%	
appropriated	10%	
tribes	07%	
Project	04%	
Budget	02%	
Conduct	01%	
Software	00%	
Complete	00%	

Figure 2: Concept Map and Rankings generated from Budget Documentation

None of the concepts seem to have much to do with software except **algorithms.** Though **Software**[5] and **algorithms** occur nearly the same number of times, **Software** is not well connected with any other concepts, having a 4% or less likelihood of co-occurring with any of the top 300 concepts in the budget and financial collection, whereas **algorithms** has a likelihood of co-occurring of up to 21% with **polarimetric,** 14% with **hyperspectral** and 13% with **model-based**. Overall, **Software** has a 7% relevance to the other content in the document collection and the concept **algorithms** has an 8% relevance. In contexts outside of this document collection, **algorithms** can be understood to have a tight meaning relation with **Software**, e.g., an algorithm is basically the logic implemented in software by software developers. However, connections between **algorithms**

[2] The automated tool used in these analyses is called Leximancer. It has been available in an evolving form for almost 10 years.

[3] Two terms co-occur if they both occur within a two sentence block. A terms consists of all word variants.

[4] Note when a term stands for a concept is in bold font and when it stands for a theme it is bold italic font.

[5] Software is capitalized whether as theme or concept because it more often appears that way than not in the budget texts.

and **Software** are almost never considered in the budget descriptions and financial summaries analyzed. The concepts are distant from one another and linked only very weakly.

Another concept related to **software** is **data**, which is both a concept and a theme in the document set. Though the theme ***data*** is not ranked as highly as ***supportability*** or ***tested***, it nevertheless has the fourth highest connectivity (46%) and relevance. The concept **data** is strongly associated with some moderately ranked concepts, like **networked** (with a 20% relevance ranking) **hyperspectral** (with a 21% ranking) and **situational**[6] (with a 13% relevance ranking) and with other concepts that are not as highly ranked like **algorithms**. For these and other reasons the theme ***data*** is therefore well integrated into the budget and financial descriptions and justifications. It also overlaps with the theme ***security***. And is somewhat associated with the concepts of **security** and **cyber**, but much less than its strongest associations.

with it). It has a connection with **infrastructure** at a moderate likelihood of co-occurrence of 9% and a connection with **networked** but at a rather low likelihood of co-occurrence (just 4%). It has some moderately strong relationships with non technical concepts like **security** (10%), **government** (10%) and **nation** (9%) but on the whole it does not mediate well between operational and technical concepts in descriptions justifying budgets.

Software as a theme is at the lowest possible connectivity occurring in the analysis, less than 1%, which registers at 0%. ***Software*** has only one concept in it, **Software**. This means that ***Software*** is an outlier theme that contains a fairly prevalent concept that is poorly integrated with the rest of the concepts in the document collection. Its distance from the other themes adds further emphasis to its outlier status (see Figure 2).

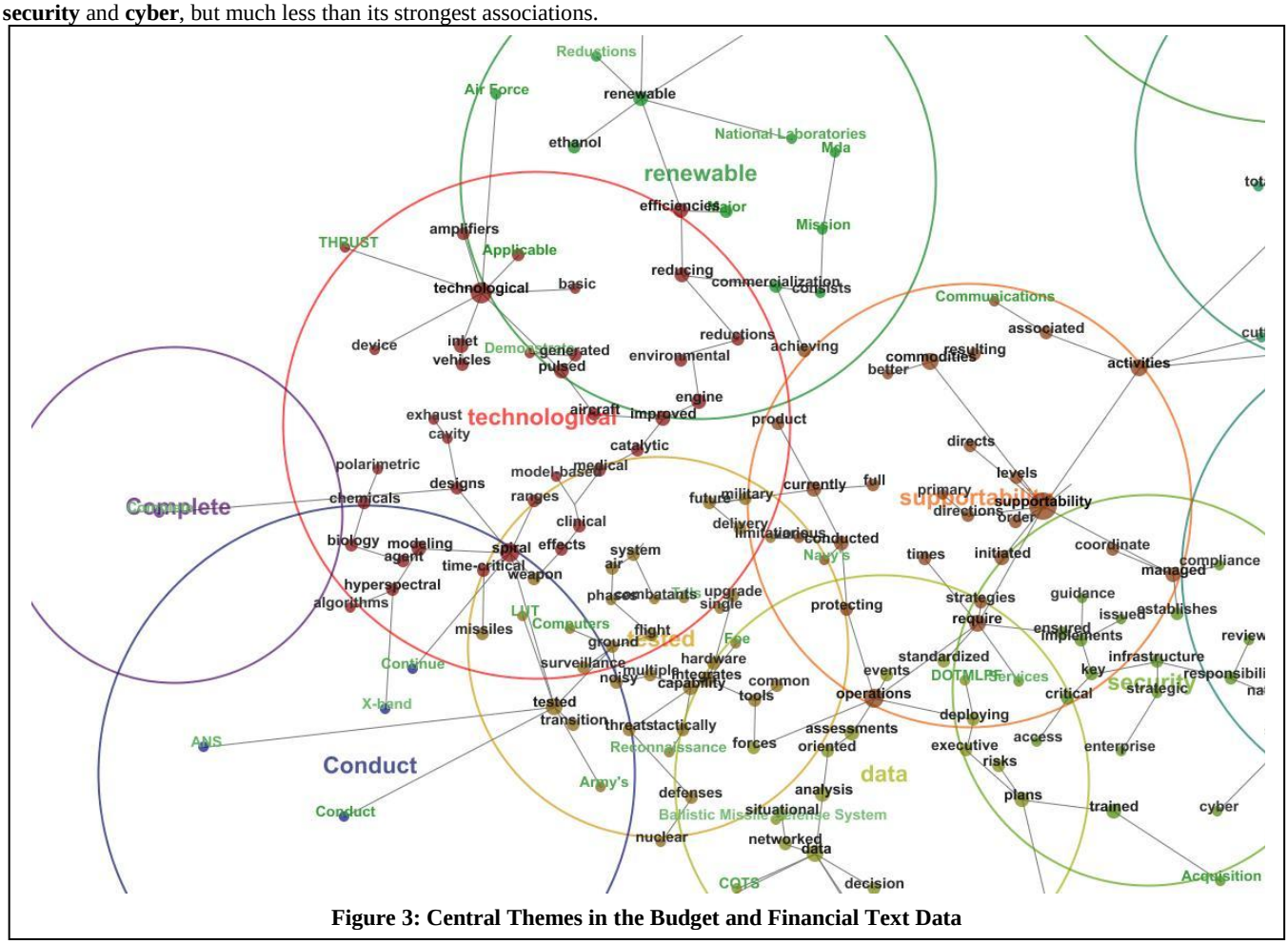


Figure 3: Central Themes in the Budget and Financial Text Data

Security is also a well-connected theme at 36% connectivity and relevance, and it contains the concept **cyber**. However, **cyber** has a near to the bottom count and relevance at 7% and is only minimally related to **Software** (1% likelihood of co-occurring

Cyber is not much better than **Software** with respect to how well it is integrated with other concepts in the collection. *It seems clear that the meaning of software as reflected in the budget rationales is impoverished and that better strategies for negotiating its meaning in all its interrelationships and complexity needs to be found.*

In addition to budget and financial descriptions and rationales, a set of documents on promising technological developments in software and cyber worlds was also collected for analysis. For this collection, **Software** is a fairly significant concept. It is ranked

[6] Situational awareness has been associated with Network Centric Warfare, which emphasizes information technology and software, but NCW has been waning in interest rather than growing after the switch from the large scale invasion of the Iraq war to small scale counter-insurgency in local situations.

third at 23% relevance of all the name-like concepts. On the other hand, it is part of a second tier theme, ***computer,*** that is just at 5% connectivity and relevance. This suggests that the concept of **Software** does not stand out from the concept of **computer** in discussions of promising technologies. What stands out more in these discussions are the themes ***security***, ranked at 100% connectivity, ***systems***, ranked at 45%, ***technology***, ranked at 32%, ***operations*** at 29% and ***Risk*** at 24%.

The last two significantly overlap with ***security***. **Cyber** is a concept in the ***operations*** theme with a rather high relevance ranking of 30%. **Cybersecurity** is a concept in the ***security*** theme but is ranked at only 9% relevance. **Cybersecurity** is a **policies** oriented concept concerned with **infrastructure** issues but with only very minimal connection to software related concepts, similarly with **cyber** (both are at 1% or less likelihood of co-occurrence with such concepts).

Perhaps a better way to mediate operational and promising technological concepts to software related concepts in this collection is through the concept of **network,** at 41% relevance and **data** at 39% relevance. Both are in the ***systems*** theme. **Data** is related to **network** at 16%, **compiler** at 14%, **dead code** at 13%, **cloud** at 12%, **World Wide Web** at 8%, **software** at 7%, **cyber** at 6% & **cybersecurity** at 5%. **Network** is related to **GIG** at 33%, **computer** at 23%, **data** at 16%, **cyber** at 16%, **World Wide Web** at 8%, **cybersecurity** at 8%, **dead code** at 4%, **cloud** at 4%, **software** at 2%. **Data** and **network** are better integrated with software related concepts than **cyber** and even provide a bridge to the latter from the former. Although **Cyber** and **cybersecurity** are more highly ranked concepts in this collection than in budget descriptions, they do not do a good job of mediating or linking operational concepts and other promising technology concepts to software related concepts. **Data** and **network** do a somewhat better mediating job, but *again better strategies for negotiating the meaning of .software related concepts are needed.*

5. Conclusion: Software Informatics and Software Engineering

Recently, software informatics has been defined as the science of information, practice, and communication around software that studies the individual, collaborative, and social aspects of software production and use, spanning multiple representations of software from design, to source code and to application [1]. One of its key characteristics is to treat software in terms of the information flows that help produce it, including design specifications, source code, documentation, software libraries, applications and services, code repositories, revision histories, and more [1]. Treating software in terms of such information flows enables it to be studied using the methods discussed in this paper, most prominently text analytics. In addition this characterization of software informatics can be broadened to include using these techniques to study and influence how users and customers of current software products understand software and what they want to do with it. To begin this work, well over ten thousand pages of governmental budget justifications for purchasing large scale typically software intensive systems were analyzed. In addition, over a 1000 pages of published literature on promising software technologies were also analyzed. We found that development, operation, and maintenance of software for these large scale systems was not much considered in the budget justifications and even marginalized in the promising technologies collection. Our basic recommendation to the ED project was to find better ways to find new ways of negotiating the meaning of software with the large varied group of software engineering stakeholders, including users and customers of large-scale software intensive systems, that were the sources of the documentation analyzed.

There are counter arguments to this approach to understanding and influencing software engineering agenda represented by the views of E. W. Dijkstra. He advanced a rather Platonic argument that progress in software engineering is not predicated on user demands but rather on mathematics (usually thought of as the pinnacle of user unfriendliness), i.e., on all the mathematical power and elegance that software can muster [3]. Dijkstra states that a *well documented* [author's emphasis] program is an object that is logically isomorphic with a constructive mathematical proof. This statement, even if not technically democratic, could very well be an hypothesis of software informatics. In other words, software informatics can study both the demand landscape of software use and software as a mathematical structure. Either way it seems to be critically relevant to software engineering.

6. REFERENCES

[1] Jones, M. C. and Twidale, M. 2009. Software Informatics? Isociety Conference, University of North Carolina at Chapel Hill.

[2] Mattern, F. Bibliometric Evaluation of Computer Science – Problems and Pitfalls. 2008. Presented at European Computer Science Summit – ECSS 2008, 9-10 Oct. 2008, Zurich.

[3] Vlissingen, R. F. van. Interview Prof. Dr. Edsger W. Dijkstra, Austin, 04-03-1985. E. W. Dijkstra Archive the manuscripts of Edsger W. Dijkstra, 1930–2002.

[4] Adler, R., Ewing, J. (Chair) and Taylor, P. Citation Statistics. Joint Committee on Quantitative Assessment of Research, A Report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICAM) and institute of Mathematical Statistics (IMS). June 2008.

[5] Mahoney, M.S. The Structures of Computation and the Mathematical Structure of Nature. 16 June 2006. paper delivered at the 21st International Workshop on the History and Philosophy of Science, The Origins and Nature of Computation, Tel Aviv and Jerusalem.

[6] Reitz, Joan M. ODLIS Online Dictionary for Library and Information Science. Special Library.

[7] Eilenberg, S. Automata, Languages, and Machines (2 vols., NY: Columbia University Press, 1974), Vol. A, xiii.

[8] Callon M, Lascoumes P, Barthes Y (2009) Acting in an uncertain world: an essay on technical democracy. The MIT Press, Cambridge (Inside Technology Series).

[9] Callon, Michel, Law, J. and Rip, A. (eds) (1985), Texts and their Powers: Mapping the Dynamics of Science and Technology, Macmillan, London.

[10] Latour, B. A Cautious Prometheus? A Few Steps Toward a Philosophy of Design (with Special Attention to Peter Sloterdijk), Keynote lecture for the Networks of Design* meeting of the Design History Society, Falmouth, Cornwall, 3rd September 2008.

Re-Gaming the Digital Divide: Broadband, MMOGs and U.S. Latinos

Julio Angel Ortiz, Ph.D.
Rutgers, the State University of New Jersey
4 Huntington Street,.
New Brunswick, NJ 08901
ortizj@rutgers.edu

ABSTRACT

Using a socio-technical theoretical lens, this paper delineates and defends the claim that access to broadband-enabled serious computer games is necessary to bridging the digital divide. Most of the research of videogames tends to focus on the potential psycho-behavioral impacts of this industry on society without regard for the unique social, cultural, and economic contexts of disenfranchised life. More specifically, little research examines how U.S. Latinos will access, consume, and use, in their daily lives, knowledge gained from serious massively multiplayer online games (MMOGs) via broadband. This void presents us with a critical moment for this research to take place and crucial theoretical and methodological problems. Latinos do not simply use videogames for entertainment or distraction. To adequately analyze and evaluate access to ICT like videogames via broadband, we must consider the larger structural and institutional forces, notably the cultural and social framework, i.e. social community informatics and social-shaping theory. Such a multidimensional perspective would help reveal the structural factors associated with Latinos' appropriation of videogames, and more interestingly how their experiences in the virtual world of videogames alters their perceptions of their real-life physical world.

Keywords

Videogames, digital divide, municipal broadband, Latinos

1. INTRODUCTION

Despite a common goal among academics, government agencies, community groups, and businesses to increase shared knowledge globally through the current revolution in technology and mass communication, a global "digital divide" persists, separating those with access to information and communication technologies (ICT) and shared knowledge from those without. The problem of access has so far been approached with an emphasis on infrastructure building. Conventional wisdom suggests that "if you build it [ICT infrastructure], they will come [unprecedented numbers of people will make use of the available shared knowledge afforded by ICT]." Prior analyses feeding the conventional wisdom, however, occurred in fields of dreams permeated with technological determinism: assuming one-way causation from technological change to social change [7, 58]. Some research suggests that the problem of access is more nuanced and complex than conventional wisdom would allow. Technologies are socially embedded and subject to the ebb and flow of cultural forces [31, 37, 42], and as such, require creative, unconventional solutions to problems of access. In this context, this study proposes a supplemental approach that could contribute significantly to bridging the digital divide—making use of the increasingly popular pastime of videogaming.

The basic premises of this research are:

1. Despite their low-brow reputation, videogames are essential to the social, economic, and political future of the United States.

2. Serious computer games, like massively multiplayer online games (MMOGs), have been evolving over recent years to the point where they now demand a broadband infrastructure.

3. Government-led interventions in the form of municipal broadband initiatives [20, 50] have proliferated recently and may prove useful in decreasing social exclusion of underserved groups in accessing the Internet and bridging the digital divide [44].

4. Little research exists on the role played by broadband-enabled serious computer MMOGs in the lives of marginalized communities, namely, U.S. Latinos

Based on these assertions, this study seeks to better understand the current and potential cultural, social, economic, and political contributions of the videogame industry to the Latino community. The researcher will examine the role of broadband in the success of this industry. The paper will also examine the ways high cost prohibits access by underrepresented groups, crippling access to the effective use of the "serious" computer MMOGs that could most benefit them. This research assumes that broadband Internet access is essential to bridging the digital divide; it also assumes government interventions can and should reconfigure access for American citizens, who, compared to other industrialized nations, face greater cost pressures for access [8, 34].

2. ACCESS TO BROADBAND IS ESSENTIAL

My approach rests on some basic assumptions about American societal values, that our society promotes the general welfare, valuing equality and fairness over elitism. While facilitating, celebrating, and rewarding individual initiative, intelligence, talent, and achievement, Americans also generally want everyone to have access to basic necessities—a living wage, food and shelter, health care, and education. The distressing pictures of societies—some of which exist within the United States today—whose members do not have access to these basic necessities provide enough grim evidence to make a good case for this argument. The main point of debate is what, within these general categories, constitutes a "basic necessity." Certain items may begin as luxuries but become so ingrained in the fabric of society that the lack thereof becomes untenable for the society as a whole as well as for individuals. It is well known, for example, most 21st-century adult Americans have a high school diploma, use a telephone, and own a car because it is universally recognized that functioning within the social order would be almost impossible without these things. Traditionally, access to knowledge and innovation has been a point of demarcation between haves and have-nots. Owning a personal computer and having access to the Internet, for example, have been, until recently, considered luxuries—certainly not basic necessities. The transition from a 20th-century industrial economy to a 21st-century knowledge economy, however, is quickly changing the standard for such consumer items as computers and broadband access (Westen 2000).

Access to and use of broadband is a fundamental part of the solution to bridging the global digital divide. Most of us take broadband Internet access for granted, and some use computer games without even thinking about it. Research suggests that broadband is not a luxury, that it is an essential component of any developed country's national infrastructure [8]. Citizens who have access to and the skills to use the Internet are: (1) more successful economically, with respect to education, jobs, and earnings; (2) more engaged politically and socially; (3) and receive more government services and other public goods than those who do not [26, 27, 34, 47, 53].

But not everyone in the U.S. is fortunate enough to have broadband access and the skills to use the Internet effectively. A large number of people in the U.S. live in rural or technologically "underserved" areas where broadband access is not available [35, 48]. Many low-income, inner-city residents simply can't afford cable or DSL-based broadband access. Public officials are aware of this need and have acted on it. The Clinton Administration, for example, championed the Internet and used the power of the federal government to encourage its growth. The Internet's rapid diffusion in the U.S. during the late 1990s was supported by a wide range of federal policies: the privatization of the Internet early in the decade; the decision to exempt online sales from federal tax; Commerce Department grants for projects that brought new communication technologies to low-income communities; and the federal "E-rate" policy of subsidizing investments in Internet technology by public schools and libraries [14, 15].

Such efforts follow a long tradition of addressing, at the federal level, universal service issues such as access to electric power, transportation, telephones, and other telecommunication services. That is, as we come to perceive that a service constitutes a "basic necessity," the federal government moves to provide access universally across the entire nation. "Universal service," then, is a policy tool [41, 44, 45]. The "digital divide," as an abstract idea, however, is something else, a political, rhetorical device. Successfully tackling the real digital divide that keeps people from accessing shared knowledge requires successful universal service strategies..

3. GOVERNMENT TELECOM INTERVENTIONS

3.1 Past Interventions

Entry into the telecommunications marketplace is coming from many directions, including cable companies (e.g. Time-Warner), wireless Internet Service Providers (EarthLink and Metro-Fi for instance) and electric utilities like Texas Utilities for example [54]. It is believed that when such entry is initiated by private companies, it contributes to the development of competition and ultimately reduces government regulation. Indeed, this is the vision of the Telecommunications Act of 1996 and the intent of the policies pursued by the FCC under the Act.

Although convergence is contributing directly to deregulation, it seems paradoxical that convergence is also luring government entities like municipal electric utilities, municipally owned cable television systems, and municipally owned wireless broadband networks into telecommunications markets [2, 4]. It appears these entities see themselves as following one of the basic tenets of "reinventing government," namely the idea that venturesome governments should find venues to participate in the marketplace in creative ways and raise capital through innovative techniques. In many ways, these governments are acting more like the private sector. Just as for private business, then, venturesome government utilities see the appearance of competition in telecom markets as opportunities for growth and expansion. Deregulation in the electricity marketplace, for instance, creates incentives for growth as government-owned electric utilities search for ways to block new entrants in their local marketplaces.

Government-owned entities have offered virtually every type of telecommunications and Internet-related service, from cable TV and local dial tone to ISP service and broadband networking. The most common communications service offered by local governments is cable television [25, 39]. A significant number of municipalities have entered the cable television market as either the exclusive provider or to compete with ILEC cable TV companies. Municipal participation in the cable TV business grew rapidly with the explosive growth of cable television during the early 1980s. However, government entrants into the telecom market hasve recently focused on delivering wireless broadband services.

3.2 Municipal Broadband Actions

In this context, responding both to a technological imperative and a lack of broadband infrastructure, nearly 400 cities in the U.S. entered the broadband market over the last five years with the intent to develop and deploy some form of municipal wireless broadband network. Government-led broadband networks are seen as a competitive alternative to other high-speed Internet services and officials argue can help bridge the digital divide, promote economic development, and/or enhance public safety for a wide range of users [20].

These initiatives continue a trajectory that has been developing for the last five years. In 2004, as newer technologies made it possible to offer wireless Internet, municipalities started entering the broadband market. Municipal leaders of this effort were creative, hopeful, and idealistic [21]. The year 2005 saw intense legislative lobbying by for-profit broadband services, resulting in a policy backlash against these first municipal entrants, unleashing a torrent of proposed state legislative restrictions. However, 2006 became the year of compromise and accommodation in which municipalities developed creative business plans so as to accommodate the needs of established service providers, as well as gain higher quality broadband service for more of its citizens. Telecom incumbents backed off from their intense lobbying efforts; as a result, a number of proposed state laws passed in a less stringent form. [4, 11]. This trend continued in 2007 and 2008 with additional municipalities entering the telecom arena, but with more and more of them outsourcing ownership and management of the service [23]. With the economic collapse in 2009, increasing numbers of municipal broadband projects were delayed or cancelled.

Although these cities initially wanted to address the digital divide and poverty with their network, the business model enforced by the providers compelled city leaders to re-think their strategy [49]. The conflict may have contributed to the failure of these projects. Most important, this reveals that, contrary to expectations, municipal networks failed to deliver. As a result, the benefits of municipal networks are exaggerated and digital divide issues are left unexplored. Given the critical importance of access to ICT in the 21st-century global knowledge economy, failure to assess municipal networks objectively and address the ongoing existence of a digital divide carries significant risks for the U.S.

4. BROADBAND GAMES

Since broadband access has major implications for our society in the years just ahead, the potential study of the topic is vast. Similarly, the impact of the videogame is comparable to that of the telephone, television, or Internet. The introduction of these earlier telecommunication devices resulted in revolutionary ways of thinking and real changes in the social fabric of American culture. To further advance our understanding of videogames as a subset media field of telecommunication, comprehend related policy issues, and develop a mature research agenda, we need to approach the subject within the context of a research program that provides economic, social, and policy perspectives.

This paper focuses on the videogame and entertainment software sector because this industry has succeeded in using broadband to connect people worldwide via online gaming and software solutions, creating a virtual community as well as sharing knowledge from virtual experiences. Despite the enormous potential of these new virtual spaces to influence, for good or for ill, our real-world societies, currently very little research exists about the social, cultural, political, and business impacts of this industry inside the U.S [12, 16, 19, 59].

Research suggests that all videogames have more than just entertainment value [9, 10]. Although categorization standards have yet to be introduced and different groups use different taxonomies, "serious" games provide educational influences on the development of social, cultural, political, and economic factors that are prevalent in today's society [5, 19, 38, 40]. Serious videogames, particularly serious massively multiplayer online games (MMOGs)[1], can be used as effective ICT tools for learning, capacity building, economic growth, and social/professional development in the 21st century. From an information technology (IT) perspective, Tapia et al argue they can also lead to the acquisition of tangible IT skills and a higher sense of self-efficacy in terms of ICT use and take-up [51]. *"Games require players to construct hypotheses, solve problems, develop strategies and learn the rules of the in-game world through trial and error. Gamers must also be able to juggle several different tasks, evaluate risks and make quick decisions… Playing games is, thus, an ideal form of preparation for the workplace of the 21st Century"* [17]. Computer games can thus serve in multiple ways to help bridge the digital divide.

The Federation of American Scientists (FAS) follows this same line of thinking. On October 17, 2006, they published the following recommendation:

> *"Groundbreaking recommendations calling on government, educators, and business to develop comprehensive strategies to use videogames to strengthen U.S. education and workforce training were released today, … America's position in the world is increasingly dependent on its standing in the technological field. Summit participants agreed that features of video and computer games can make learning more effective and accessible by teaching players higher-order learning skills."* [18]

Unfortunately, providing access to the cyber infrastructure for our population is apparently fundamentally different than providing services such as telephones and electricity [49, 50].

[1] *World of Warcraft* and *Linage* are good examples of serious MMOGs

Despite efforts over a decade to provide all segments of the U. S. population with low-cost broadband access, a digital divide exists. As such, I believe that universal access to broadband-enabled MMOGs can be a key component in reaching some of the social, cultural, political, and economic goals of our society.

5. A THEORETICAL FOUNDATION FOR THE BROADBAND + GAMES EQUATION

In this section, I argue that ICTs are neither a sufficient nor a necessary condition for ushering in an era of universal access. However, it is also evident that ICTs like broadband and videogames primarily driven by commercial interests are here to stay. It is therefore urgent that a socio-technical approach that uses these vehicles be better understood in our information society. The social vision proposed rests on four key themes: 1) Going beyond connectivity; 2) Ensuring capability components, 3) Promoting content; and 4) Creating context-specific environments. In the vision proposed, broadband-enabled MMOGs are not inherently necessary or beneficial. The challenge is, precisely, to be able to tell when, and under what conditions, these ICTs can contribute to development for marginalized communities in the information society

5.1 ICTs are not Neutral

Theoretically, information technologies are not neutral. According to some studies, information technologies embody the values of particular industrial civilizations where technical mastery defines the dominance of one group over another [28]. Technologies are fashioned by social groups to promote their values [43]. To examine the digital divide from a critical perspective will enable the researcher to discuss underlying assumptions that inform cultural and technological inequality. Examining the digital divide with the goal of bridging it gives voice to those with far less power in the public discourse on technology in our rapidly changing society. It also poses new research questions that challenge and test the limits of long-held assumptions, such as the conventional wisdom about access to and ability to use technology or how new ICT are separate from any specific context or cultural understanding.

5.2 The Information Society

Classic works on the topic describe the information society as a meritocracy that enables a level playing field [6]. This is certainly the goal of bridging the digital divide. Expected efficiencies garnered from the integration of network technology, telephones, and computers can potentially revolutionize all industries, including finance, manufacturing, and advertising. New technologies certainly offer several potential advantages, including: stimulating economic development, facilitating electronic commerce, enhancing educational pedagogy, improving health care in remote locations, and providing an impetus for electronic democracy.

In Bell's theoretical meritocracy, individuals from underrepresented groups can effectively compete with wealthy and privileged individuals. However, before this meritocracy is achieved, factors such as limited educational and employment opportunities must be addressed. Broad patterns of social inequality in education, work, and consumption opportunities are at the heart of the digital divide. The digital divide reflects an ongoing social inequality in the U.S. that can be explained by both a lack of vision and entrenched social, economic, and political systems [14]. Social inequality shapes diffusion rates and the rate of IT use. This unequal IT usage is a reflection of existing social inequalities. [15, 29, 30, 31].

5.3 Universal Service Theorized

The universal service obligation is a cornerstone of American industrial and regulatory policies. It is probably the major building block of the concept of public service which is central to regulatory policies. Historically, universal service has typically been provided by a monopolistic public or regulated operator and its financing mechanism has been designed accordingly. The goal of universal telephone service has never been simple to define. Research has focused on what exactly this goal is and how it might be achieved [41]. Most policies surrounding universal service are strongly oriented toward the infrastructural aspects of telephony. This focus is reflected in the measures used for universal service: telephone penetration rates, whereby availability of service is described as physical access to the telephone network.

As universal service moves into the information age, its context has forever changed, yet its message is still the same. In essence, universal service is about guaranteeing communication ubiquity, both within the home and beyond. However, universal service carries a good deal of telephone-related conceptual "baggage" that needs unpacking if it is to be a useful principle for the Internet. If this is not addressed, universal service will continue to be a 1930s solution to a 21st century problem. Measuring it in the information age requires a research design capable of identifying the societal effect of policy and impact of organizational/municipal design over time. Universal service needs to be forward looking to help build the broadband networks of the future in a manner that fulfills the vision of the Constitution.

6. LATINOS AS POTENTIAL BRIDGERS

Because the digital divide is more than digital and a symptom of deeper, more important divides [3, 33, 46, 56], a discussion of the digital divide engenders discussions concerning pre-existing socio-economic and racial disparities. It is in this regard that special attention should be given, if any significant impact is to be made in mitigating this social ill, to the largest and fastest growing minority population in the U.S., Latinos.

Hispanics (i.e. Latinos) are rapidly changing the cultural face of the United States. This growing demographic has presented challenges and opportunities to program planners, policymakers, business leaders and service providers. The U.S. Hispanic community has grown from 22.4 million in 1990 to 35.3 million in 2000, a 58% increase – over four times the total growth of the U.S. population. Studies project that by 2050, Hispanics will not only form the majority of the white population and surpass the African-American population's rate of growth [22], but account for approximately 30% of the population compared to 14% in 2005 [36]. Today, research already shows the majority of first graders in the top ten U.S. cities are Latino (Tomas Rivera Policy Institute 2008). It is undoubtedly the fastest growing major race or ethnic group in the American population [24, 54, 55]. By these statistics alone, it is unequivocally clear that Hispanics will impact the nation just as profoundly as Irish, Italian, and Jewish immigrant populations who influenced and dominated the changing American culture of the 19th and 20th centuries.

This said, however, Latinos currently comprise a minority that is not fully participating in the technological revolutions. They are on the wrong side of a digital divide that if not rectified could adversely impact this population—and consequently American society. Because videogames are more popular with Latinos than traditional Internet access and more conventional uses of the Internet [36], this research is predicated on the possibility of using videogames to help this group cross the digital divide. In order to do so, this research project will assess the impact of broadband-enabled MMOGs on Latinos.

According to the Pew Internet, videogame playing is almost universal among teenagers across all groups. Apart from cell phone videogames, there is no difference in the devices used by teenagers to play videogames. Lower income teenagers, however, are more likely to say that they play cell phone videogames than upper middle class families. This is interesting in light of the fact that teens from upper income (+$75k per annum) families are only marginally (79% vs. 63%) more likely to own cell phones [36].

Race and ethnicity seem to influence the types of games played. Black and Latino teens are more likely to play sports, adventure, fighting, and survival horror games than their white counterparts, who are more likely to play more complex online games. MMOG are most popular among white youth. Since many of the advantages in terms of training and teamwork come through MMOGs, it should be noted that MMOGs are also more popular among teenagers from educated households. Households with parents whose education background is high school or lower tend to be more likely to play videogames alone [36].

The significance of these statistics speaks to the role of group dynamics in videogames. Although there are arguably some advantages to solo play, MMOGs are still more advantageous from a societal perspective. Research has shown that MMOG players learn important business and social skills through their multiplayer environments. These skills include working well with others, maintaining a social setting with highly volatile group dynamics, leadership, division of labor, and communication [1, 13]. If a significant number of players are white and upper middle class, the social status quo is merely perpetuated. Inducing underprivileged classes like Hispanics to play these games will prove a significant step toward bridging the digital divide and creating a meritocracy where everyone can compete on a level playing field in the global digital economy.
.

7. CONCLUSION

I have argued that we must establish a deeper understanding of the social, cultural, economic, and political impacts of the videogame industry in the U.S, specifically of MMOGs in relation to U.S. Latinos. Specifically, we need to understand how a computer game–driven MMOG broadband model may boost digital inclusion and thus mitigate social ills engendered by information inequality;

Given the pivotal role of broadband access in the information society, further research is still required to explore the role MMOGs play in the lives of U.S. Latinos and how their experiences in the virtual world of videogames, via broadband access, alter their perceptions of the real-life physical world (e.g. family, work, religion, etc.). Similarly, future research may suggest ways in which governmental actions can support infrastructure that promotes serious computer MMOGs via broadband by at-risk communities.

It is important to note that progress has been made in telecom and municipal broadband reform. This said, nonetheless, the most difficult challenge of policy implementation remains on the horizon. Establishing effective and responsive federal, state and local legislation that furthers our free-market enterprise all the while fulfilling the growing needs of all consumers will be a challenge. Specifically, testing the limits of competition in unbalanced markets and creating laws that promote economic development and universal access will require extremely informed and very competent policy makers. The daunting task of providing universal access must be prioritized on governmental agendas if the increasing divide between the telecom/knowledge and disparaged groups is to ever be fully addressed. As federal and state legislation takes center stage in this new decade, the legislations being crafted reflect this increased dependency on ICTs like municipal broadband as information networks and serious videogames as knowledge drivers. The opportunities are enormous and the challenges are unlike those encountered in the past history of telecommunications.

8. REFERENCES

[1] Aldrich, C. (2004). *Simulations and the future of learning*. San Francisco: Pfeiffer.

[2] Allen, S. (1985). *Cable Television: Municipal Use of government access, 1972-1982*: University of Maryland

[3] Babb, SF. (1998). The Internet as a Tool for Creating Economic Opportunity for Individuals and Families. *Dissertation*. University of California at Los Angeles, Los Angeles, CA.

[4] Bar, F and Park, N. (2006). "Municipal Wi-Fi Networks: The Goals, Practices, and Policy Implications of the U.S. Case," *Communication & Strategies*, **61**(1), 107-125.

[5] Barab, S. A. (2003). Designing for Virtual Communities in the Service of Learning. *Information Society* 19(3), 1-7.

[6] Bell, D. (1974). *The Coming of Post-Industrial Society: A Venture in Social Forecasting*. London: Heinemann.

[7] Bijker, WE. (1995). "Sociohistorical Technology Studies." In S. Jasanoff, GE Markle, JS Petersen & T Pinch (eds.), *Handbook of Science and Technology*, (pp. 229-256). Thousand Oaks, CA: Sage.

[8] Bleha, T. (2005). "Down to the Wire," *Foreign Affairs*, May/June 2005.

[9] Block, Deborah (2009). Video Games Promote Peace and Democracy. *Voice of America*. 21 September 2009. Accessed on September 25, 2009 from http://www.voanews.com/english/2009-09-21-voa37.cfm.

[10] Bryce, J and Rutter, J (2002a). Computer and Video Gaming: *Academic Perspectives, Positions and Research Resources*. CRIC discussion paper, University of Manchester, Manchester UK

[11] Christensen, A. (2006). "'Wi-Fi'ght Them When You Can Join Them? How the Philadelphia Compromise May Have Saved Municipally-owned Telecommunications Services," *Federal Communication Law Journal*. June 2006.

[12] Crandall, Robert W. and Sidak, J. Gregory (2006). Video Games: Serious Business for America's Economy. *Entertainment Software Association Report*, 2006.

[13] DeMarco, M., Lesser, E., O'Driscoll, T. (2007). Leadership in a Distributed World: Lessons from Online Gaming. *IBM Institute for Business Value Report.*

[14] DiMaggio, P, Celeste, C, and Shafer, S. (2004). "Digital Inequality: From Unequal Access to Differentiated Use." In K. Neckerman (ed.), *Social Inequality*. New York, NY: Russel SAGE Foundation.

[15] DiMaggio, P, Hargittai, E, Neuman, W, and Robinson, J. (2001). "Social Implications of the Internet," *Annual Review of Sociology*, 27, 307-336.

[16] Dyer-Witheford et al (2005). "The Political Economy of Canada's Video and Computer Game Industry." *Canadian Journal of Communication*. **30** (2).

[17] The Economist (2005) *Defending video games: Breeding evil?* Accessed on July 5, 2009 from http://www.economist.com/opinion/displaystory.cfm?story_id=4247084. August 4.

[18] Federation of American Scientists (2006). *National Summit on Educational Games*. Accessed on July 5, 2009 from http://www.fas.org/programs/ltp/policy_and_publications/summit/index.html October 17.

[19] Gee, J. P. (2005). Why video games are good for your soul. *Pleasure and learning*. Melbourne: Common Ground.

[20] Gillett, SE. (2006). "Municipal Wireless Broadband: Hype or Harbinger*?" Southern California Law Review*, 79, 561- 594.

[21] Gillett, SE., Lehr, WH., and Osorio, C. (2004). "Local Government Broadband Initiatives," *Telecommunications Policy*, 28, 537-558.

[22] Guzman, B, The Hispanic Population: Census 2000 Brief, in c2KBR/01-3, U.S. Dept. of Commerce (ed.). 2001, U.S. Census Bureau.

[23] Hauge, J, Jamison, M, and Gentry, R. (2007). "Bureaucrats s Entrepreneurs: Do Municipal Telecommunications Providers Hinder Private Entrepreneurs? " *Information Economics and Policy*. **20**(1), 89-102.

[24] Hernandez, JC. (2000). "Understanding the Retention of Latino College Students," *Journal of College Student Development*, 41, 575–588.

[25] Jacobson, R. E. (1977). *Municipal Controls of Cable Communications*. New York: Praeger.

[26] Katz, J and Rice, R. (2002). *Social Consequences of Internet Use: Access, Involvement, and Interaction*. Cambridge, MA: MIT Press.

[27] Kennard, W. (2001). "Equality in the Information Age." In B. Compaine (ed.), *The Digital Divide: Facing a Crisis or Creating a Myth*. Cambridge, MA: The MIT Press.

[28] Kling, R and Lamb, R. (2000). "IT and Organizational Change in Digital Economies: A Sociotechnical Approach." In E. Brynjolfsson & B. Kahin (eds.), *Understanding the Digital Economy* (pp. 295-324). Cambridge, MA: The MIT Press.

[29] Kling, R, McKim, G, and King, A. (2002). "A Bit More to It: Scholarly Communication Forums as Socio-Technical Interaction Networks," *Journal of the American Society for Information Science and Technology*, **54**(1), 47-67.

[30] Kvasny, L. (2002). Problematizing the Digital Divide: Cultural and Social Reproduction in a Community Technology Initiative. *Dissertation*. Georgia State University, Atlanta, GA.

[31] Kvasny, L and Payton, FC. (2005). "Minorities and the Digital Divide." In M. Khosrow-Pour (ed.), *Encyclopedia of Information Science and Technology*, (pp. 1955-1959). Hershey: Idea Group.

[32] Lenard, TL. (2004). "Government Entry Into the Telecom Business: Are the Benefits Commensurate With the Costs?" Progress on Point, Release 11.3 February 2004, *Periodic Commentaries on the Policy Debate - The Progress and Freedom Foundation*. http://www.pff.org/issues-pubs/pops/pop11.3govtownership.pdf

[33] Norris, P. (2001). *Digital Divide? Civic Engagement, Information Poverty and the Internet in Democratic Societies*. New York: Cambridge Univ. Press.

[34] Oden, M. (2004). *Beyond the Digital Access Divide, Developing Meaningful Measures of Information and Communications Technology Gaps*. Austin, Texas: The University of Texas at Austin Press.

[35] Oden, M and Strover, S. (2002). *Links to the Future: Information and Telecommunications Technology and Economic Development in the Appalachian Region*. Washington, DC: Appalachian Regional Commission.

[36] Pew Internet and American Life Project (2008). *Teens, Video Games, and Civics: Teens' Gaming Experiences Are Diverse and Include Significant Social Interaction and Civic Engagement*. Accessed August 20, 2009 from http://www.pewinternet.org/Reports/2008/Teens-Video-Games-and-Civics.aspx

[37] Pinch, T and Bijker, W. (1987). "The Social Construction of Facts and Artifacts." In W. Bijker, T. Hughes and T. Pinch (eds.), *The Social Construction of Technological Systems* (Vol. 17-50).

[38] Pursel, B. K. & Bailey, K. D. (2007), *Establishing Virtual Learning Worlds.* Accessed on August 20, 2009 from http://www.virtualleamingworlds.com/vlw_working.pdf

[39] Rizzuto, R. J., & Wirh, M. O. (1998). *Costs, Benefits, and Long-Term Sustainability of Municipal Cable* Television Overbuilds: Tele-Communications Inc.

[40] Quinn, C., & Conner, M. (2005). *Engaging learning: Designing e-learning simulation games*. San Francisco. CA: John Wiley &. Sons

[41] Sawhney, H and Jayakar, K. (2005). Universal Access: Precedents, Prevarications and Progress. Paper presented at the *Telecommunications Policy Research Conference* (TPRC), Alexandria, VA.

[42] Sawyer, S and Tapia, A. (2005). "The Sociotechnical Nature of Mobile Computing Work: Evidence from a Study of Policing in the United States." *International Journal of Technology and Human Interaction*, **1**(3), 1-14.

[43] Shields, M. (1997). "Reinventing Technology in Social Theory. " In *Current Perspective in Social Theory*, Vol. 17, pp. 187-216. Greenwich, CT: JAI Press.

[44] Schement, JR. (1999). "Of Gaps by Which Democracy We Measure." *Information Impacts Magazine*, December 1999, www.cisp.org/imp/december_99/1299schement.html.

[45] Schement, JR and Curtis, T. (1997). *Tendencies and Tensions of the Information Age: The Production and Consumption of Information in the United States*. New Brunswick, NJ: Transaction Publications.

[46] Selwyn, N, Gorard, S, and Williams, S. (2001). "Digital Divide or Digital Opportunity? The Role of Technology in Overcoming Social Exclusion in U.S. Education." *Educational Policy*, **15**(2), 258-277.

[47] Servon, L. (2002). *Bridging the Digital Divide: Technology, Community and Public Policy*. Oxford, UK: Blackwell.

[48] Strover, S, Chapman, G and Waters, J. (2004). "Beyond Community Network and CTCs: Access, Development and Public Policy." Telecommunications Policy, 28, 465-485.

[49] Tapia, AH, Powell, A, Ortiz, JA. (2009). "Reforming Policy to Promote Local Broadband Networks." *Journal of Communication Inquiry*. **33**(4), 354-375.

[50] Tapia, A and Ortiz, J. (2006). "Municipal Responses to State-Level Broadband Internet Policy." Paper presented at *The 34th Research Conference on Communication Information and Telecommunications Policy Research Conference* (TPRC), Alexandria, VA.

[51] Tapia, A, Seif El-Nasr, M, Yucel, I, Zupko, J, Maldonado, E. (2007) "Building Virtual Spaces: Games as Gatekeepers for the IT Workforce." *International Federation of Information Professionals (IFIP) Working Group 8.2/9.5.* July. Portland, OR. (presenter)

[52] Thomas, M. L. (2004). Government Entry Into the Telecom Business: Are the Benefits Commensurate With the Costs? *The Progress & Freedom Foundation.* Feb. 2004.

[53] Tufekcioglu, Z. (2003). *In Search of Lost Jobs: The Rhetoric and Practice of Computer Skills Training*. Austin, TX: University of Texas at Austin Press.

[54] U.S. Census Bureau. (2000). *Projections of the Resident Populations by Race, Hispanic Origin, and Nativity: Middle Series, 2050 to 2070 (NP-T5G).* Accessed on March 4, 2006, from www.census.gov/population/projections/nation/summary/np-t5-g.txt

[55] U.S. Census Bureau. (2003). *Income, Poverty, and Health Insurance Coverage in the United States.* Accessed on September 3, 2009 from www.census.gov/hhes/www/income.html

[56] Van Dijk, J and Hacker, K. (2003). "The Digital Divide as a Complex and Dynamic Phenomenon." Paper presented at the *Proceedings of the International Communication Association*, Acapulco, Mexico.

[57] Westen, T. (2000). E-Democracy: Ready or Not, Here It Comes. *National Civic Review*, **89**(3).

[58] Winner, L. (1986). "*Do Artifacts Have Politics?" The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: The University of Chicago Press, pp. 19-39. Accessed on January 3, 2005 from http://www.courses.psu.edu/phil/phil403_pam208/winner/index.html

[59] Wolfe, Joseph & Crookall, David (1998). "Developing a Scientific Knowledge of Simulation/Gaming," *Simulation & Gaming*, Vol. 29, March, 7-19.

Navigating the Social Terrain with Google Latitude

Xinru Page, Alfred Kobsa
Donald Bren School of Information
and Computer Sciences
University of California, Irvine
Irvine, CA 92617 USA

xpage@uci.edu

ABSTRACT

Although researchers have been building location-based social services for some time now, sharing one's location has only recently been introduced to the more general population. This paper examines real-world adoption of and resistance to Google Latitude, a social mobile-device application for people to share their locations. We report findings from an analysis of semi-structured interviews with 21 participants using grounded theory. Our research reveals how interviewees perceive the social affordances of location-sharing applications to be conceptually intertwined with the conventions of other social networking and communication technologies; Our findings emphasize that many participants felt pressured to not only adopt social applications such as location-sharing, but also to be responsive and accessible at all times. Participants perceived technology-mediated social interactions (such as "friending" someone) as highly symbolic, and as problematic if they did not strictly adhere to the established social etiquette. We also found that participants' perception of the social norms around using Latitude varied widely, affecting how and whether participants used the system.

Categories and Subject Descriptors

K.4.2 [**Computing Milieux**]: Computers and Society – *Social Issues.* H.5.2 **[Information Systems]**: Information interfaces and presentation – *User Interfaces: Evaluation/methodology.*

General Terms

Human Factors.

Keywords

Google Latitude, Location-based services, Facebook, Twitter, Instant Messaging, Social Networking, Adoption, Social Norms, Symbolism, Mobile Devices

1. INTRODUCTION

In February 2009, Google launched Latitude, a real-time location-sharing feature within Google Maps that has been adopted by over a million mobile phone and laptop users [17]. Also gaining popularity are other location-sharing products such as Loopt, GyPSii, Whrrl and Foursquare. Researchers point to the benefits of disclosing location in cell phone conversations, including creating social or process awareness, coordinating meetings, and signaling availability, caring, or need for help [3]. On the other hand, though, concerns have been raised about stalkers, abusive spouses and a panoptic Big Brother. Heated arguments over these issues as well as cautionary media reaction reinforce the importance of understanding people's attitudes towards location-tracking and their adoption or rejection of this technology. This is specifically in regards to real-time location-sharing that broadcasts location continuously or frequently.

Our previous research showed that people's attitudes towards Google Latitude are deeply connected to their use of other social networking technologies [13]. In fact, they perceive location-sharing social applications as an additional type of social networking technology, and hence its use cannot be studied in isolation. Since recent studies point to the psychological benefits and social capital gained by participating in social networking [9], it is vital to investigate the reasons why people do and do not participate. However, little is known about who is and is not using social network technologies [6], let alone social location sharing. Thus, this report describes people's real world attitudes towards adopting social location-sharing applications, with many findings also being relevant to other social networking technologies.

Our analysis reveals that the most salient factors surrounding adoption are social influences (both real and imagined) and not so much the popularized privacy or security issues. We identified the following three factors: 1) social pressure to use this technology, 2) the symbolic meaning behind technology-mediated social interactions, and 3) users' understanding of the social etiquette surrounding technology use.

2. Previous Research

Much location-tracking research within the location-based services literature emphasizes privacy concerns. By probing hypothetical scenarios via questionnaires, experiments and experience sampling methods, researchers found that people's willingness to disclose their location depends largely on who is requesting it, and also why [11, 8]. However, stated privacy attitudes often differ from actual behavior [18]. Thus, a few studies looked at location-tracking usage of prototypes [10], including real-time disclosure with small social units (e.g. group of friends) who volunteered to use it [15, 1, 7, 19]. These studies showed that location awareness facilitates coordinating meetings, checking on loved ones, and social connectedness. However, studies of location tracking within real world social connections have been few.

Sociological studies on online technology use in people's pre-existing social networks revealed that usage is shaped by preexisting motivations [5]. However, theories that characterize modern day social networks (e.g. Wellman's "individualized networking" theory) still lack empirical validation [5] and may not extend to social networking technology. Many "personal network" studies look at only a subset of ego's complete network (e.g., the strongest relationships) [20], due to the cost and difficulty in generating a complete social network. In researching how people use social networking technology, it is crucial to consider weak relationships that outnumber strong ones in friend lists [9].

This research has been supported by NSF Grants IIS 0808783 and 0953071.

Some ethnographic studies focus on social communication technologies such as instant messaging [12], but little research has looked at technologies that convey locational presence information rather than mainly serving communication purposes.

3. Methods and Sample

This report is based on semi-structured interviews with 21 individuals, conducted mostly one-on-one and face-to-face (participants beyond driving distance were phone interviewed, and a husband and wife pair was only available for a joint interview). Informed by theories of innovation diffusion [16], framing [2], privacy [14], and trust [4], we crafted open-ended questions to ask about their experiences with Latitude, their feelings towards using it with various contacts and contexts, and about alternative ways in which they connect with others. Because little is known about who is and who is not using social networking technologies [6], we recruited 10 interviewees who had not used Latitude and 11 interviewees who had used it. Since Latitude was new and likely to have attracted the attention of those more technically inclined, we recruited participants through student discussion lists in Information and Computer Sciences at UC Irvine, through non-academic personal contacts from various locations in the United States, and through subsequent snowball sampling.

The interviewees consisted of 4 females and 17 males with ages ranging from 21 to 40's (averaging 28). Of the 10 interviewees who had not used Latitude, 7 had decided not to use it and 3 wanted to but did not own a supported device. Of the 11 interviewees who had used Latitude, 7 were still using it and 4 had abandoned it. In terms of other social technologies, all but one interviewee used Facebook or Orkut, instant messenger was similarly popular, and about a third used Twitter. With regard to their relationship status, 13 were single, 2 living with a significant other, 1 in a long distance relationship, and 5 married with children. Their professions ranged from graduate student (some having previously worked in industry), software developer, product marketing manager, lawyer, and construction project manager, to housewife. 15 participants were either born in the United States or had lived here for five or more years. 6 participants were originally from Asia (mainly India) and had been here one year or less.

4. Results

Oftentimes reported attitudes were very similar across participants. Sometimes though, a bifurcation arose between about half of participants who were largely optimistic about using Latitude to improve their lives, and the other half who believed location sharing would be a burden. In this paper we refer to the former as the *optimists* and the latter as the *pessimists*. What surprised us is that this division in attitudes does not align with participants' usage of Latitude: some were reluctant to use Latitude but still tried it out, while others were optimistic about it but had abandoned it. In the remainder of this paper we will therefore distinguish between the *optimists* and *pessimists* rather than partition interviewees by their use of Latitude. We believe an understanding of the underlying motivations and attitudes with regard to Latitude sheds more light on the adoption question than do micro-level reasons for using or not using these technologies.

We analyzed our interview transcripts using grounded theory and open coding. In this paper, we report on three significant themes: social pressure and technology addiction, symbolic importance of social interactions, and understanding and construction of social etiquette surrounding location-tracking technology. We note that differences in attitudes between different genders, cultures, occupations or fields of study, age groups, and relationship status were not as poignant to these three themes, but were relevant for themes that we will discuss in a future paper.

4.1 Latitude: The next CrackBerry?

Optimists often used Latitude just because it was new, and couldn't wait for a critical mass of other adopters to really make use of the technology. "Tirtha", a graduate student using Latitude on his laptop, lamented, "I send so many invites, and nobody's responding to it, so it's like feeling 'aw... nobody is going to see me, why should I update [my location]'". *Pessimists* were divided: There were those like "Chris", formerly the product marketing manager of a major tech company, who wants "there to be some critical mass" so as not to "waste my time on sorting out the weaknesses of" new technologies. On the other hand, there were those who dreaded reaching a critical mass: "I feel like it's where we're headed. There's enough people that will just say yes to all of it – to Twitter to Latitude," begrudged "Elizabeth", a graduate student.

Once a technology hits critical mass, interviewees felt compelled to stay on it. As "Jared", who had often bought the latest gadgets when he was in industry, recounted: "I started on twitter mostly because it was new and I wanted to try it out. Same with Facebook. And then everyone got on it, so you stay on it." Elizabeth elaborated on the social pressure to stay on: "Twittering is more like the kind of pressure I might feel if people were using Latitude. Or Facebook. It's like if you don't exist…on Facebook, you don't exist in this place and you're not part of this place… there's all kinds of questions why doesn't this person exist."

Both *optimists* and *pessimists* felt pressure to convey their location responsibly. "Jake", a graduate student, expressed his dismay at finding that his friend was not using Latitude to represent his true location: "…if I see a person toy with an application, I just won't pay attention to them on it. So it's like my confidence in how well they use it. So once it's broken, I'm not going to pay attention after that." Similarly, another student "Lee" lamented how his perception of Latitude changed because it allows users to manually set their location; Lee believed this meant that "it's often not accurate [and] becomes this irrelevant piece of trivia. This is what someone says where they are, not actual... It reminds me of Twitter if you were just to get random Twitter posts from a bunch of people and put it on a page."

Furthermore, interviewees felt the pressure to engage fully in the technology. One interviewee compared this responsibility to how she recently started using a calendaring application and "now it's like if I forgot this [meeting], it's a greater slight than it was before when I just used to be late all the time." "Dan", a software developer, describes his resistance to Latitude as the same as for instant messengers: "Somebody can always look and see whether I'm online or not… IM expects you to respond immediately if they know you're online." Interviewees even described themselves as poor users when they did not keep location or status up-to-date or check their wall posts often enough.

Interviewees often used the word "addiction". Chris was actively working on not crossing the thin line from BlackBerry user to CrackBerry addict (as Chris describes it, someone who feels compelled to constantly use the technology). Dan described how

his teenage son could not turn off his phone despite tiring of the constant texts:

> "Sometimes he'd just put his phone to the side and walk away from it. He got tired of responding to text and he felt like he couldn't turn it off for some reason. It had to be on, so he wouldn't miss a text but he just didn't want to do it. He was having friends text in the middle of the night, so he was having trouble sleeping. So when he lost his phone, I think that he decided that it was a good time to take a break."

Many *pessimists* were worried about being sucked into a new technology that they would have to maintain and engage in fully. Some even avoided Latitude in consideration of others. "Noah", a construction project manager asserted: "If I didn't always want them to know where I am, I'm not comfortable with always knowing where they are. Or I wouldn't want to walk in with the same issues with them - I wouldn't want to walk in on… [their] business meeting, in a family gathering, or whatever." "Ankur", a graduate student, put it more strongly: "I'm kind of banking if they want to get in touch with me, they'll call me themselves. I wouldn't want to force somebody to meet me."

Other *pessimists* used new social networking and communication technologies but fought against social pressures by minimizing participation and maintaining a wallflower-like online presence. Lee limited his Facebook posts to others' walls in order to slow down friends' posting to his. He even stopped using status in instant messenger because it invited friends to interrupt him.

4.2 Friending as a Handshake

The importance of social interactions in Latitude and other social technologies rests in their symbolic meanings. By and large, interviewees went to great lengths in order not to offend others, including changing their own technology use and behavior.

The most common example of changing behavior was that of friending someone on Facebook or Latitude, i.e. requesting someone to accept you as a friend in their friend list. Elizabeth explained her first experience with receiving a friend request: "my gut reaction was that it would be bad to not accept… it's like a handshake. Friending someone is like putting out your hand, and saying no is like not shaking their hand… And now you're in this long, you've started this thing where you are now friends. And now you have to deal with all the stuff that that is." The symbolic gesture of accepting a friend request is so strong that participants overwhelmingly reported an eclectic mix of contacts in their friend lists, dominated by weak ties.

Furthermore, in social networking technologies these weak ties are symbolically on par with strong ties. They're accorded the same privileges and require upkeep. "Derin", a graduate student using Latitude, illustrated this by pointing out that even though "you can choose people to see your location in the city level, but probably they can understand you're sharing your location at the city level because it's not accurate. You don't want people to see that you're not sharing your information... It's kind of rude. It will basically make them question your relationship or friendship."

However, giving the appearance of being friends on technologies like Facebook was often a sufficient symbolic gesture. Many interviewees granted partial profile access to weak ties in a way that those ties could not visually differentiate it from full access. For the most part this sufficed. However, interviewees occasionally ran into problems when an acquaintance would overstep his bounds and try to use restricted features. As a result, some interviewees resorted to the lesser of two evils and instead ignored friend requests. Others went with "the lowest common denominator" of disclosure, i.e. they disclosed merely that information about themselves that everyone may see. Still others elevated the privileges of weak ties and put up with the consequences:

> "I mean it's funny, I actually get annoyed by [happy birthday posts] because people…write you a quick note, and then now I feel like…I have to respond [and] I have to make every single one unique because I don't want to just paste one thing for everyone and make it look like a thoughtless person. I just feel like it introduces this load of work which is totally unnecessary and just has no real end."

While these practices around friend etiquette were prevalent, married participants in our sample were an exception; they were not as concerned about the symbolic meanings of friending, although they still did engage in lowest common denominator disclosure.

Another symbolic interpretation involved ones' willingness to use Latitude. Interviewees commonly equated using Latitude as a statement of trust, or of having nothing to hide. Ravi explained, "I'm thinking, as the common man, if someone is skeptical of using that device...what's the problem he's facing that he doesn't want to use the technology. So I'm not a criminal, and I don't have those sort of feelings." Elizabeth further related her significant other's unsuccessful attempts for them to use Latitude together: "it became an issue of trust. [Bobby] said to me, 'why wouldn't you let me. I wouldn't abuse it.'" "Fei", a corporate lawyer, also recognized the symbolic significance for her relationship. But when she installed Latitude in the company of someone she had just started dating, she did not feel compelled to friend him: "I think there was a mutual understanding that we didn't want to know where each other were all the time. Like we weren't in that phase of our relationship." Interestingly, none of these interviewees were concerned about the symbolic meaning of using or not using Latitude with their closest relationships, but many dreaded acquaintances and superiors who may want to connect. "Eric", a computer programmer, bemoaned, "That would be a really hard decision. I'd probably just add them just based on the fact that they're my manager."

Furthermore, some interviewees also considered a request to use Latitude as symbolizing that the requester wants you to know the minutia of his or her real-time location. This seemed egotistical for those beyond very close friends or family. Sam limited sharing his location to a small group of techie friends so he wouldn't appear "egotistical". *Pessimists* complained of being overwhelmed by others egotistically sending out status and minutia from Facebook, Twitter and even instant messaging status. Chris complained, "People abuse Twitter and Facebook…I want to know how friends are, but I don't want to know that they're at the movies with their son, right? It's like, status update doesn't mean I want to know exactly what you're doing at all times of every day…Latitude's got the same problem." This egoism caused many *pessimists* to avoid new technologies such as Latitude.

4.3 The Salad Fork Goes Where?

Lastly, we discuss interviewees' understanding of social etiquette. Some *pessimists* expressed uncertainty and angst over the social etiquette of new technologies. This manifested as a lack of knowledge about what others were doing. Fei explained how she stopped using Latitude as soon as it asked her to add friends:

> "I wasn't sure what would happen after I started adding friends. If it would be weird. I didn't really know the etiquette with Latitude. It's sort of like twitter. I didn't really know how to use twitter. I know from a technical perspective. But I just don't know the etiquette of it… I wouldn't have known when it would become weird for like an acquaintance-level friend to get this type of invitation from me. Because I've never gotten an invitation to join from anyone."

Elizabeth expressed uncertainty over whether what she saw was the same as what others were seeing:

> "I don't really know what everyone's doing. Because I don't know what other people see because it's not reciprocal. If somebody else's settings are different, so if I make a judgment based on the norms I see, it's so inflected by my settings. If it turns out that everyone these days who signs up for Facebook goes in and sets all the parameters in a certain way because that's the sort of etiquette to set, then I'm already seeing a weird version of Facebook because of my setting. I don't know it's so complicated."

While some were searching for this social imaginary, other *pessimists* drew from past experience with social networking technology to project norms in the new technology. Depending on the type of technology from which they drew, and their individualized experience of it, this understanding of the social norm differed from person to person. This often led to angst that the projected social etiquette would have undesirable consequences.

Optimists on the other hand, often confidently conveyed what they imagined to be the social norms. Many based this understanding on their own actual or projected behaviors. However, we saw that these behaviors differed from person to person. For example, Jared explained the norm of Facebook stalking:

> "so you know there's a newsfeed; it's yelling at you in an Internet sense. It's kind of stalkerish but everyone does it so it's not so weird. Yeah, but that's how you keep track of, lets me know what my friends are doing up north more than other mediums. Or actually rather than talking to them, hey what are you doing this week, and you can see, oh he went to the cherry blossom festival, and here's the photos, and here's the video."

In stark contrast, Jake asserted:

> "So then I go and see how their page is updated. Although I don't look too much, because then I feel creepy looking around at their pictures. So I'll just look at some random stuff. Oh they're working at this company now. Oh so and so wrote on their page. There's something creepy about going in and looking at a whole bunch of their pictures for hours… it just feels weird. They gave their Facebook. And so the way I use it is just go through some pictures of stuff that's happened. And so I realize that's how they use it too. And so just because they gave me their Facebook, it feels kind of weird looking into all the stuff that they did without actually asking them about it."

With all of these various conceptions of social etiquette and norms, there was not a single common social norm for any of these technologies. Rather, individuals each came up with their own understandings and acted upon it.

5. Conclusion

We found that various social pressures guided people's adoption decisions for social location-sharing applications as well as for other social networking technologies, namely:

1) Once technologies gained a critical mass, interviewees felt social pressure to join. Moreover, once they joined they felt obligated to fully participate, being accessible and responsive. This led to some *pessimists* resisting social pressure to use Latitude and other technologies.
2) The symbolic meaning of friending someone and of using Latitude greatly shaped how participants use these technologies, sometimes leading people to outright reject them.
3) Interviewees had different understandings of the social etiquette surrounding these technologies. *Pessimists* were either uncertain about norms or drew from their past experiences of other technologies to project norms. *Optimists* often derived norms based on their own behavior, which varied considerably from person to person.

We recommend that developers of location-based social services consider addressing these social pressures in order to attract a broader audience and help existing users participate in social networking more fully. We plan to continue studying social location-sharing technology adoption as it gains popularity, including the continued impact of social influences on its use. We are also developing design proposals for addressing people's concerns with Latitude by drawing on other findings from this study.

6. ACKNOWLEDGMENTS

Our thanks to Bonnie Nardi and Sameer Patil for their feedback on prior drafts.

7. REFERENCES

[1] L. Barkhuus et al., "From awareness to repartee: sharing location within social groups," Proc. of the twenty-sixth annual SIGCHI conf. on human factors in computing systems, Florence, Italy: ACM, 2008, pp. 497-506.

[2] R.D. Benford and D.A. Snow, "Framing Processes and Social Movements: An Overview and Assessment," Annual Review of Sociology, vol. 26, 2000, pp. 611-639.

[3] F. Bentley and C. Metcalf, "Location and activity sharing in everyday mobile communication," CHI '08 extended abstracts on human factors in computing systems, Florence, Italy: ACM, 2008, pp. 2453-2462.

[4] G.A. Bigley and J.L. Pearce, "Straining for Shared Meaning in Organization Science: Problems of Trust and Distrust," The Academy of Management Review, vol. 23, Jul. 1998, pp. 405-421.

[5] J. Boase and B. Wellman, "Personal Relationships: On and Off the Internet," in D. Perlman & A.L. Vangelisti eds.

Handbook of Personal Relations, Cambridge: University Press, 2006.

[6] D.M. Boyd and N.B. Ellison, “Social Network Sites: Definition, History, and Scholarship,” Journal of Computer-Mediated Communication, vol. 13, 2007, pp. 210-230.

[7] B. Brown et al., “Locating Family Values: A Field Trial of the Whereabouts Clock,” Proceedings of UbiComp 2007, Innsbruck, Austria: Springer, 2007, pp. 354-371.

[8] Consolvo et al., “Location disclosure to social relations: why, when, & what people want to share,” Proc. of the SIGCHI conference on human factors in computing systems, Portland, OR: 2005, pp. 81-90.

[9] N.B. Ellison, C. Steinfield, and C. Lampe, “The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites,” Journal of Computer-Mediated Communication, vol. 12, 2007, pp. 1143-1168.

[10] G. Iachello, I. Smith, S. Consolvo, M. Chen, and G.D. Abowd, “Developing privacy guidelines for social location disclosure applications and services,” Proceedings of the 2005 symposium on usable privacy and security, Pittsburgh, Pennsylvania: ACM, 2005, pp. 65-76.

[11] S. Lederer, J. Mankoff, and A.K. Dey, “Who wants to know what when? Privacy preference determinants in ubiquitous computing,” CHI '03 extended abstracts, ACM, 2003, pp. 724-725.

[12] B. Nardi, “Beyond Bandwidth: Dimensions of Connection in Interpersonal Communication,” Computer Supported Cooperative Work, vol. 14, 2005, pp. 91-130.

[13] X. Page, A. Kobsa, “The Circles of Latitude: Adoption and Usage of Location Tracking in Online Social Networking,” IEEE International Conference on Computational Science and Engineering, Vancouver, Canada: 2009, pp. 1027-1030, DOI 10.1109/ CSE.2009.195.

[14] L. Palen and P. Dourish, “Unpacking "Privacy" for a Networked World,” Proc. CHI 2003, 2003, pp. 129-136.

[15] M. Raento and A. Oulasvirta, “Designing for privacy and self-presentation in social awareness,” Personal and Ubiquitous Computing, vol. 12, 2008, pp. 527-542.

[16] E.M. Rogers, Diffusion of Innovations, Free Press, 2003.

[17] M. Siegler, “While people worry about Facebook photos, a million users let Google know exactly where they are,” VentureBeat, Feb. 2009.

[18] S. Spiekermann, J. Grossklags, and B. Berendt, “E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior,” Proceedings of the 3rd ACM conference on Electronic Commerce, Tampa, Florida, USA: ACM, 2001, pp. 38-47.

[19] R. Want, A. Hopper, V. Falcão, and J. Gibbons, “The active badge location system,” ACM Transactions on Information Systerms, vol. 10, 1992, pp. 91-102.

[20] B. Wellman, “The network is personal: Introduction to a special issue of Social Networks,” Social Networks, vol. 29, Jul. 2007, pp. 349-356.

Of mouse and men: Computers and geeks as cinematic icons in the age of ICTD

Joyojeet Pal
ATLAS Institute, University of Colorado at Boulder,
TASCHA, University of Washington, Seattle
4311 11th Ave NE, Seattle WA 98105
510.501.8679
joyojeet@uw.edu

ABSTRACT
Since the early 2000s, Information and Communications Technology for Development (ICTD) has rapidly gained attention as an emerging area of scholarship as information schools increasingly diversify their interests into issues of socio-technical systems and technology adoption issues in the developing world. Research in ICTD while vibrant, has been largely restricted to issues around the use of information technologies towards a range of activities including eGovernance, computer-based learning, agricultural information systems. In this article, we turn to the under-researched discourse of technology in the developing world to probing at the a rather heretical questions of why people living in various forms of deprivation find hope for economic and social development in technology. Our starting point in this research is the outcomes of 196 interviews among rural Indians with no primary experience with technology, but a great deal of enthusiasm about using or training their children to use computers. We found that this enthusiasm about technology was primarily based on secondary sources of information, a large part of which was cinematic representation of computers and computer users in local movies. Investigating this in popular Indian film, we find a visible positive and highly aspirational discourse of technology both in the representation of technology users and the artifacts themselves, such as laptops or the internet, a trend particularly evident on comparison with western cinema. To discuss the issue of intentionality in this trend, we interview leading filmmakers in India and find that unconscious absorption of social aspiration into the scripting, and significant intent into the use of computers and computer users as symbols of modernity that filmmakers feel Indian audiences respond positively to. We propose that the use of India as a case for broader examination is important on two levels. First, regionally, India is a 'leader' in the ICTD movement both because of the symbolic value of its software sector despite the co-existing underdevelopment and also because it is home to a range of ICT-based initiatives aimed at bringing about developmental outcomes. Secondly, from the perspective of the future of ICTD within information studies, the role of media, especially popular film, in the construction of knowledge about technology is an important and under-researched area that this article seeks to take forth.

Keywords
ICTD, India, Cinema

1. INTRODUCTION
Much recent scholarship has examined the optimistic discourse of technology as means of social inclusion throughout several parts of the developing world. Images of children in impoverished surroundings tapping away in front of flatscreen monitors or holding up spiffy laptops have adorned television screens, billboards, and academic journals alike, as a vast number of technology companies, international agencies, governments and non-profits have invested in products or programs meant to provide technology to play a role in easing communication, learning, capacity-building or administrative barriers to development. US-based information schools have played an important leadership role, both on the technology and social science scholarship around 'Information and Communications Technology and Development (ICTD).[1] The involvement of multiple stakeholders and rapid rate of increase in both programs and projects in this space has meant that scholarly focus on ICTD issues has mostly been restricted to issues of design, technology adoption, and project evaluation to understand impacts. In the process, there has been little assessment of the evolving discourse of technology in this regard, especially in light the relatively unimpressive record in achieving developmental goals, for many of the programs implemented in this space.

This work emerged as a by-product of work in four districts of rural south India, where researchers interviewed 196 residents to document ideas about technology among people with no experience using computers. Results showed a role played by secondary information sources such as mass media in helping build what had come to be a strong positive view of technology and its possibilities in discussions of the rural residents speaking of their own futures. Discussions revealed that the trusted institutional channels of information on technology were popular media and electoral speeches, both of which we thereafter found taking broadly optimistic positions on technology and its value in society.

In this study, we focus primarily the portrayal of computers and computer users in Indian movies. Building on past scholarship in technology and development, cinematic portrayals, and the role of icons in aspiration, we argue using films since the 2000s that popular cinema is an important reflection of the prevalent discourse of technology in India. There are three trends we outline – first the artifact of the computer itself and its portrayal as symbolizing power. Second, we record a high number of screen

[1] Other acronyms including ICT4D, ITD etc.

protagonists playing technologists, and find that the portrayal of their occupation is generally very positive and aspirational. Finally, we find that this trend in popular cinema is much more concentrated in south India. We frame these findings through issues of gender, class, and geography within the recent history of the development discourse in India.

Finally, while the centrality of popular cinema in building national discourses of aspiration is possibly unique in several aspects to nature of media consumption in India, we argue that conceptions of development both as defined by a society for oneself as well as for an "other" are deeply influenced by reflections of aspiration in popular media.

2. RELATED WORK

We are concerned in this article with the construction of the identity of computer users and of artifacts such as the computer and the internet in popular Indian cinema, and how this in turn is a reflection of aspiration in the information age. Identifying popular cinema as a mode of institutional information, we situate this work as contributing to recent work in film theory both broadly within the sphere of nationhood and aspiration [1-3]. In our discussion of technology and development, we are interested in issues of identity in transition, thus past work on identity on celluloid post-fascism and dictatorship in Germany [4], Spain [5], Latin America [6] as well as on issues of identity within a globalizing context, thus work on transnational Chinese identity [7] or value negotiation on issues of religion of "African-ness" in Ghanaian or Nigerian video films [8, 9]. We draw on a rich body of work on Indian popular cinema oriented both more broadly on issues of national identity [10] and more specifically on cultural change [11, 12].[2]

Issues of new media have been enthusiastically embraced in Information Studies, including those of representation and identity online [13, 14]. On issues of development, there has been work on the creation of online identities of Diaspora communities of Filipinos [15] Chinese [16] and Nigerians [17], for instance, or on issues of self-representation in Sub-Saharan Africa [18]. The negotiation and careful crafting of online identity and networks has also been studied with regard to groups such the Zapatistas [19] and West African online-scamsters [20]. However, an important exceptions of this work has been the lack of significant interest in the impact of the virtual identity on individual aspiration, an area we delve into in this paper. Interestingly, while iSchools have enthusiastically taken on the realm of identity and aspiration in an environment of greater connectedness within the new media space, there has been little attention to the construction of aspiration through existing popular media portrayals in the information age.

Finally, we draw from a growing literature on ICTD in information studies, engineering, and in development studies. ICTD is arguably rooted in work on the evolving nature of economic and social power relations based on networks [21, 22], and more broadly on an emergent 'information society' [23-25], all of which point to a reconfiguration of economic and social relations across classes and between nation-states in this era. ICTD has been examined from the perspective of social inclusion [26], stakeholder theory [27] and the technology artifacts [28] themselves. Particularly relevant to the direction we seek has been literature in this space on identity formation and articulation [29-31], transnationalism [32, 33], , and gender studies [34, 35].

3. METHODOLOGY

We reviewed 91 films, of which 47 films are specifically Indian films from the 1990s and 2000s which depict computer users or artifacts in some form. These films are largely in the Tamil and Telugu languages, for reasons we discuss later. A textual analysis of the films helps us make broad generalizations about underlying meaning of the films, but leave unanswered the question of intentionality in the narrative. To explore this, we spent time in the film industry discussing the ideas emergent in this article and interviewed 5 people from the Tamil and Telugu film industry whose work is discussed here, these include 4 directors and 1 distributor: Rajiv Menon, Suhasini Mani Ratnam, Siva Anantasubramanian, Siddharth, and Dr. Srinivasan from Abirami Films, the largest distribution house in South India. Each of the professionals interviewed in this sample are associated with what are referred to as A-list films, typically very large budget popular films. Given the difficulty of access to these filmmakers and the expense involved in identifying makers of these films and interviewing them, we were able to conduct only 5 interviews, which are primarily used to reflect on the central themes identified in the analysis of the films.

4. DISCUSSION

Ideas of technology and society have dated back over a century to the fanciful silent short film *Trip to the Moon* [36], and the representation of technology in cinema has straddled the line between science fiction, and what may be, and technology and social readjustment. The latter, especially the idea of technology as transformative, came to centerstage in the 1920s and 30s, around the period and often theme of rapid industrialization. Scholars have been interested in two important and fairly consistent themes along technology and society in cinema – the first, the dystopian ideas of technology and urban living – either in the present or future as seen in Fritz Lang's futuristic epic *Metropolis* set in a fractured 21st century or Charles Chaplin's *Modern Times* [37], set in a dehumanized 20th century factory floor. These ideas of technology as mystical, all pervasive, and potentially at odds with humanity have been a consistent theme of cinema and literature throughout the decades of vast technological change around the world [38, 39]. A second early theme had a more proactive view of technology, primarily industrial production, and bears its origins in the early Soviet cinema highlighting technology as a nationalist enterprise. This theme, often attributed to Lenin's view of cinema as a means of social and economic propaganda [40] was later emulated several other nationalist cinematic traditions around the world, where cinema has been seen as both a symbol of and a simultaneous propagandist tool of modernity [41]. Broadly under this umbrella have been the two related themes – the first was that of a struggle between traditionalism and modernism, one seen several 'third world' cinemas [42, 43], and second was of the use of technology (especially big industry) in nation-building, which India in

[2] An area of recent growth has been in activist cinema, which appeals to widespread audience, such as Abderrahmane Sissako's Bamako (2006), in addition to copious documentary cinema on the subject.

particular, saw a fair share of in the post-colonial years [44].

The first films to feature computers were science newsreels in the immediate postwar period [45]. The rising popularity of television in the US spurred the appearance of computers on the small screen, starting with the 1962 sci-fi show *The Jetsons* [46], in scattered episodes of the spy caper serial *The Avengers* [47], and finally in 1966 with the hugely popular sci-fi series *Star Trek* [48]. The transposition of the Martian-equivalent unknown to the fantastic machine was an automatic next theme in the dystopian imagination of computers guided by the popular conceptions of artificial intelligence – as typified by the man v/s machine face-off in *2001: A Space Odyssey* [49]. Computers in general remained restricted to the mad-scientist or spy caper space, thus reducing their meaning in the "everyday" context [50, 51]. The idea of a typical workplace scenario invaded by computers remained a relatively low-impact area, with the rare exception of films like the 'library sciences classic' *Desk Set* [52] in which Katherine Hepburn plays a librarian who risks being rendered jobless by a computer. The individual human intermediary of computers on screen remained by and large the socially-awkward scientist. It wasn't until the late 1970s and early 1980s, when the use of computers in schools increased significantly, and media introduced us to the young geek, who would come to typify the computer user in films such as *Hide and Seek* [53], *War Games* [54]. The ubiquitization of computers on screen, where the computer moved out of being a scientific spotlight to being an everyday item started around the late 80s with films focusing on industries where computers had become fairly commonplace, such as banking in *Wall Street* [55], and by the 1990s, there was a huge spike in technology-related blockbuster cinema made for popular consumption in Hollywood, around the time of the Silicon Valley economic boom, [56] and especially the widespread permeation of the Internet.[3]

The imagery of progress has been a central theme of Indian cinema in portrayal of the tension between tradition and development through most of its post-colonial history. A rich body of work has examined the representation of modernity in cinema [57-59], and more recently on television in the post-liberalization India [60]. In looking at the way computers have been represented on film, we find many of these themes repeated including man-machine dystopia [61, 62], the mad-genius scientist or supervillain with electronic den [63, 64], or the science-fiction fantasy [65-68]. However, two factors stand out as unique in the Indian case. First, simplistically expressed ideas of computers as fantastic machines have virtually disappeared from western cinema, though this remains a common theme in many Indian films. Second, the computer engineer as an aspirational hero seems like an outlandish lead character in western cinema, whereas a programmer protagonist is an extremely common theme in many of the new south Indian films.

[3] This is somewhat comparable to the appearance of computers and computer users in India, which became much more common a decade later when home computers achieved reasonable (though far from ubiquitous) permeation in urban India.

4.1 Regional Calibration

Our analysis here is largely restricted to films of South India. The commonly used term 'Bollywood' typically refers to the Bombay industry that makes Hindi film, the derivatives 'Tollywood' and 'Kollywood' refer to the Telugu and Tamil[4] cinema. The emphasis on Tamil and Telugu language films are in part due to the foundational data from respondents in the earlier study on information sources about computers, but more critically because a comparatively greater trend in the unique depiction of computers in aspirational terms on screen in these films than in Hindi cinema.

"We think about every smallest detail in the characterization. It is essential for the common viewer to digest the protagonist's profession. It may be different for other parts of the country, but if you present a software engineer as a hero, even a villager in Andhra Pradesh will immediately pick it up. It will not be considered elitist."

Sivakumar, Director (Telugu)

One of the first major south Indian films to feature a computer-centric theme was the 1986 bond-esque action flick starring Kamal Haasan, Vikram [69]. The film, written by popular science-fiction author Sujatha had a caper plot, and the casting of Kamal as the star was important at the time since the actor was himself emerging as an urbane star, from the shadows of several decades of rural-themed filmmaking. In subsequent years, there would be an occasional film with a protagonist using computers [70], but this was typically not central to the theme of the film until Tamil director Mani Ratnam's 1992 blockbuster hit, Roja [71], in which the protagonist is a cryptographer. In many ways, Roja was a landmark film within this genre for several reasons – first, the protagonist was a south Indian, living among north Indians, second, the desirability of a computer-engineer groom was central to the theme of the film, and third, the protagonist was not an omnipotent macho man, but rather a soft-spoken, romantic, righteous hero.

Roja was a rare film that did well both in its native language, Tamil, and in north Indian Hindi-speaking markets, bringing with it a new stereotype of a south Indian software engineer. While the theme of computer engineers was not entirely absent in Hindi films, we argue that these followed a narrative description fairly comparable to western cinema in that the character used technology if needed, but it never quite became a desirable quality in a man. In subsequent years, the Hindi film engineer characters typically came across as 'cool youngster' hacker characters, usually teen idols, [72-75], very comparable to the Hollywood depictions of computer teenager users from the 1980s and 1990s whereas computer-users in South Indian films, as we discuss in subsequent sections, tended much more to be aspirational heros. We attribute this distinction to two factors. First, this finding supports existing work in film theory that has discussed a 'westernizing trend' of Hindi cinema [59], with a greater upper class urbane narrative, and marketed to urban elites, thus films

[4] Kollywood because of the studios' location in the Kodambakam neighborhood in Chennai, the capital of Tamil Nadu.

distanced aimed at an audience that switches between Hollywood and Bollywood with equal ease.

The second, and more compelling reason is the sheer composition of the technology industry, which has had a far greater south Indian component. Furthermore, south Indians in software industry were not just residents of metropolitan cities like Chennai, Hyderabad or Bangalore, but rather came from smaller towns and villages, thus consumers of regional language films rather than the metropolitan tastes which are relatively more Hindi-language oriented. From the films sampled, the greatest concentration of software-engineer characters is in Telugu cinema, which is explained from the fact that smaller cities in Andhra Pradesh such as Vijaywada and Visakhapatnam were among the biggest contributors of young engineers to the software boom in India.

Related to this second reason is the architectural transformation of the South Indian city and its relevance to cinema. In older films, it would be common in films to explore the urban-rural divide through images of village folk coming to cities and marveling at 10-storey buildings or flyovers. Following the 1990s and the plethora of new construction in these cities, the exploration of the same themes of inequity or social change was conveniently done through the frame of the technology industry. As opposed to the past when films about the small town migrant to the city would be spiced with disaster, the reality of increasing labor market flexibility in the South for educated workers would increasingly be reflected on screen.

4.2 Aspiration

"If we want to show a modern scene for the audience, we can either rent a Mercedes car, or show a café with some young people and a few laptops. This is cheaper from the art direction perspective, and shows youth, modernization, technology – all in one."

Siddharth, Director and head of a film marketing company

Take for instance the following plot from Mani Ratnam's subsequent urban Tamil superhit, Alai Payuthey [76]. In this, the protagonist Karthik is an engineering student in love with Sakhti a medical student. The class-crossed couple meets on the commute to college; the two fall hopelessly in love, decide to marry, to which the arranged-marriage inclined parents throw both out of their respective homes.

This kind of face-off between youthful love and parental opposition to marriage is a time-tested theme of Indian cinema. Back in the old days, Karthik would probably have been reduced to begging office to office for a desk job wearing a tie, finding in his useless paper degrees a metaphorical foil for the oppressive market economy [77] and thereafter turn to a life of dubious ethical distinction [78, 79]. Sakhti meanwhile would probably sit home sacrificing square meals and running a bare-bones household with a sanctimonious smile [80]. A number of themes, including the dependence of the man on the system to earn an honest living, the helplessness of the woman outside of the home domain, and the importance of parental consent and wisdom would typically be highlighted in the couple's struggle.

Instead, in *Alai Payuthey*, the couple turns to a new direction for its salvation – technology. Karthik starts a computer software company with his friends, eventually winning an outsourced contract from the US that fixes for good their financial troubles, rubbing in the process a few parental noses in the dirt on the gold-paved streets of south Indian cities. Karthik fails most stereotypes of cinematic occupational characterization. He is no idealistic teacher, nor upstanding cop, nor charismatic businessman. He is at best a lovable nerd, traveling on a scooter to work daily with a laptop strapped to his back instead of a holster. In effect, he is the archetype of exactly what he isn't in the movie – an educated gentile -- the perfect candidate for an arranged marriage.

"It has become a compulsion to plug computers into the movies. Most of the people who wrote films in the past had biases against educated heroes, as the profile of those who make and those who watch films has changed, so have the characters."

Sivakumar, Director

Following the early lead of Roja's cryptographer groom, two broad strands of software engineer-related marriage scenarios have emerged – the first in which the engineer is the middle-class hero [81-83] and the second is the counterpoint – in which parents are shown hankering after a groom who is a software engineer or NRI, and the hero in this case is usually a son-of-the-soil type [84-86], and while the endgame of these films is often an ode to the anti-hero, the focus on the software engineer as essential to middle class aspiration is nonetheless highlighted. In the blockbuster Telugu hit AMAV - Adavari Matalaku Ardhalu Verule [87], the protagonist is a good for nothing who cannot speak English and can't get a job, much to the chagrin of his father. Eventually, he turns his life around by taking a job in a software company, becomes a star programmer, learns English, wins a woman of his choice, but yet remains traditional in his values by dutifully handing over his paycheck to his proud father each month. The film was so successful that it was remade in every other south Indian language by a local star of that state.

Figure 1: An exotic location, swimwear, and an urbane superstar, with a MacBook on the publicity poster for Kandasamy(2009)

In some films, the less than desirable character turns to computers as a means of social acceptability and in others [88, 89], the drive is primarily economic, often explicitly as a means of getting jobs in the US [90, 91]. The relationship with immigration is also an interesting theme within aspiration, stemming from the fact that a fairly significant number of educated engineers from southern states had opportunities to work or study abroad, especially in the

US. The 2008 Telugu hit film Chintakayala Ravi [92] entirely revolves around a village-migrant protagonist who is a bartender in the US, but whose family believes he has a job as a software engineer and proudly proclaim the same to all and sundry through the movie. The film turns into a class tension study as a marriage proposal becomes the plot turning point when the highly desirable software engineer groom turns out to be a "lowly" alcohol server.

4.3 Iconization

"Computer can save us. When neighbouring country is attacking, this is known to our scientists by tracking it on the computers."

Udhaykumar, 5th grader, Coimbatore, Tamil Nadu

The quote from Shivraj at the start of this was not an isolated one. We were surprised at how often in our interviews in rural India, the same answer was repeated to us over and over "Computers can do anything" sometimes ranging from children's fantasies to adults with no direct experience with computers allocating human attributes to the machines, "Computers can teach us English" in clear seriousness (in the film AMAV discussed above, this is virtually true) It is far-fetched to ascribe such notions specifically to films, but it's worth looking briefly at some of the omnipotent deeds of computers. In the previous section, the movies we describe are perhaps a degree of sophistication above a second strand of films in south India – the mass films, which rely heavily and very successfully on an iconic cult of personality. Such films are written and marketed around a actor rather than the plot.

There are two elements to the iconization that are relevant in such cinema – the user and the computer as an iconic artifact itself. While much work has focused on the protagonists and their characterizations as elements of modernity [93-95], little has focused on the actual artifacts themselves. The computer as an iconic device, often portrayed as something magical may seem to be a simplistic description, appealing to the naïveté of the audience, especially when juxtaposed against the discourse of fanciful enthusiasm about computers as we see from the youth above. What we refer here to as simplistic views of computers could on one hand include exaggerated descriptions of an off-the shelf computer's abilities, such as unusually advanced voice recognition in *Sivaji - the Boss*[5] [96], computer programs can estimate with remarkable accuracy what a child will look like when they grow, auto-adjusting for sartorial grace and facial pounds in *Dharmapuri* or *Vaitheeshwaran* [97, 98].

The point with these films is not so much that the audience is expected to swallow the magic of the machine, but more so that the machine is part of the iconic male hero that commands it. Thus, the real question here is how well the machine fits in as an accessory to an omnipotent star.

"There is no star with a bigger draw than Rajnikant. For his fans, he is perfect, he is their leader. If there is a latest technology, Rajnikant should be able to use, it in the eyes of the viewers."

Suhasini Mani Ratnam, Director

[5] 'BOSS' in the film stands for Bachelor of Social Science

The iconization of the screen actor is an important element of the image politics of south Indian cinema, and several of the major film stars are state legislators or political bosses. Such actors typically tend to play larger-than-life characters in what is referred to as 'mass' cinema in trade circles, films that have widespread appeal among the rural and urban poor [93]. Most elements of narrative around the star in such film has a closely controlled populism, bordering on propaganda, and can probably be traced back to the cinema of M.G.Ramachandran, who never drank or smoked on screen, committed no acts of questionable morality with a screen image closest to the righteous Lord Rama of Hindu mythology. Like MGR, several of his populist actor successors go by honorifics such as "Dear Leader" "Ultimate Star" "The Captain" and "Young General." In each of the four south Indian states of Karnataka, Tamil Nadu, Andhra Pradesh, and Kerala, there are at least a couple of very successful movie stars that fit the category of such larger-than-life stars with frenzied fans.

Figure 2: Actor Vijaykanth in a political poster guised as various national heroes

Thus as opposed to the truer-to life depictions of aspirational heroes in some of the films described in the previous section, such 'mass stars' do not stray from their typical characterizations, since scriptwriters are constrained by maintaining the image of the star. Thus a typical mass star plays roles that bring him closer in identification to the proletariat, as opposed to something that could be construe as elitist such as, and their use of the computer is more strategic as if to indicate that despite identifying with the people on class terms, the character can master and use technology as needed.

"If we write a film for urban audiences, we avoid even the slightest error – for instance, we had a film with a US-returned engineer character, and at a wedding proposal, he mentions the name of his college and the department he was in. It turned out, we had the department name wrong, and the director chastised the staff for this. For the kind of film that is oriented around a big star with a mass rural following, the scriptwriter would not pay this kind of attention. In the shot, the star would use the computer, and that is enough explanation."

Sivakumar, Director

Thus in a range of films featuring such mass appeal stars, the use of the computer is subsumed within the screen character's role, typically fighting evil in society. One of the most popular such uses of a computer has been in maintaining databases of villains, that the hero will proceed to eliminate with very emphatic strikethroughs on a monitor or some such visual confirmation, as used by megastars Vijaykanth [99], Chiranjeevi [100], and Ajithkumar [101, 102]. A variant of this has been films featuring online websites maintained by the hero to solicit citizen complaints against corrupt government of anti-social elements that the star thereafter uses to swiftly provide justice to the wronged [103, 104].

Although a great number of such portrayal may seem anywhere between comical and fantastic, the functional value of the overall association of political persona with technology clearly holds meaning within an environment of economic liberalization and the technology sector in particular finding a place in political populism in south India [105]. In the past decade there has been work on the association of technology in crafting and managing public images for politicians icons in south India, such as with Chandrababu Naidu [106] and SM Krishna [107], former chief ministers of Andhra Pradesh and Karnataka.

4.4 Gender

In the hugely popular Tamil adaptation of Sense and Sensibility, *Kandukondain Kandukondain* [108], the female protagonist Sowmya, a Tamil Brahman version of Jane Austen's Elinor Dashwood takes on the responsibility of reversing her family's economic woes, and as a first step towards this, leaves the village with her family in tow. She moves from her village teaching job to the city and becomes a software engineer, excels at work, and eventually gets offered a position in California. The characterization of Sowmya contrasts with stereotypical female characters written for Tamil films in three important ways – economically, she takes on the role of a man, geographically, she opts for the city, and professionally, she drops a 'woman's job' of teaching and turns techie. We interviewed director Rajiv Menon to understand if there was intentionality in this description.

"When the software movement started, it was emancipator – for example, Jane Austen wrote Sense and Sensibility before the Suffragist movement, so in my adaptation, the two protagonists stand for art and knowledge. The heroine moves from a rural to urban setting and out of poverty by becoming a software engineer. My protagonist was not to be an angelic face of rural ethic. I saw this as a meritocracy."

Rajiv Menon, Director

Sowmya's character was by no means the first independent, driven female character in Tamil cinema (and that she quits her plans to work abroad in the last scene to marry the male lead is somewhat dampening). What is distinct about the character in contrast to strong female protagonists in the past is the lack of serious social opposition to the steps she takes. Kandukondain was in many ways perfectly emblematic of the growing openness to females in the workforce, including in villages where women moved out to cities, and this was clearly emphasized several times in our interviews. During our field research in rural Karnataka in 2005, we met a 21-year old female teacher at a village computer center. Both her parents were illiterate laborers.

"I want to move out of the village. I am looking for a job with computers because my parents will let me move to Udupi or even Bangalore if the work is in computers. For any other job, they won't let me leave the village."

Geetha, Computer Teacher, Udupi, Karnataka

About a year later, an interesting corollary to Geetha's statement came from a taxi driver interviewee, in Tamil Nadu, a father of two girls in their 20s.

"Both my daughters work in Chennai in computers. In the early days, we would never let our (referring to the Thevar caste) women travel to the city to work, but if they work for computers that is good. There are good facilities with only ladies housing, and many other families from our village have sent their daughters to work in Chennai now."

Selvaraghavan, Taxi Driver, Coimbatore District, Tamil Nadu

These anecdotes are from feeder towns where many of the female workers at call centers of Bangalore and Chennai originate. A body of work around Indian call centers that probes issues of gender and empowerment (or lack thereof) in the urban migration experience [109-111] has been growing rapidly. This phenomenon plays awkwardly into Indian cinema. If we think back to women on screen, "normal" women didn't have careers, or even jobs. And those that did, typically did so because of the failure of some critical male provider. Thus, the dead, drunk or incapacitated rural husband or father gives way to the woman who works the field, at the risk of lascivious attention of the agrarian landlords or plantation managers [112-114].

Which brings us to the second theme related to Kandukondain – migration. When a female character is required to give up the safety and virtuosity of the village for a city to seek employment, her typical fate would be a nasty exposure to the corruption of urban life. The theme of urban life as being threatening to an unprotected woman's sexuality is consistent across regional cinema in India [115, 116]. Typically, the lumpen exposure to urban workplace lechery came through the stereotypical wicked building contractor [117, 118].

In general, a screen female was safest when she worked in "mother figure" such as nurses or doctors [119-121] or caring social workers or teachers [122-125], reasonably respectable jobs for women, and not macho enough for the male protagonists. The female character who spills over into the organized employment sector almost certainly does so out of need. The villainy of the rural landlord and the urban building contractor are now reprised in the lascivious white-collar bourgeois rogue harassing female co workers [126-128], with perhaps the only exceptions to the rule being the rare daring saleswoman [129-131] or a spoilt heiress boss [132, 133]. In most of these cases, a male character, usually the hero, offsets the perils of the woman's tryst with the man's domain of the economy either in the form of a benevolent boss or a kind co-worker. There is the occasional vengeance-themed film where the female protagonist must take on the role of Hindu

goddess Durga as a police woman [134, 135] or avenging angel [136, 137].

Other than these, the woman's screen job degenerates quickly in the caste hierarchy of professions from the mildly uncomfortable bar dancer or performer [138, 139], to the circumstantial prostitute [140-142], and finally to the campy gangster's moll [139, 143, 144]

The association of technology jobs as being desirable is a particularly unique trend, since this cinematic legitimization of women in technology is reflected in discussions with respondents in villages who spoke approvingly of female relatives or others in social contacts who lived in cities working in call centers with computers. While there had probably never been an Indian film with a female character playing an engineer save for the oddball automobile mechanic's daughter, the lack of physical contamination or visual masculinity in technology jobs allowed a convenient blend into the accepted image of womanhood in film and we see a burst of female software engineers on screen [145-151], including films with a reversal of roles – where the female lead plays an accomplished technologist of some form, and the male lead is portrayed as professionally subservient [152, 153].

Figure 3: Actress Nayantara wears a software company lanyard, as a headstrong engineer in Yarodi Nee Mohini (2008)

While a majority of the portrayals are within the safe zones of not ruffling expected depictions of female leads, the boundaries are occasionally pushed In the Telugu film *Anand* [154], the protagonist Roopa takes Sowmya's position a step further by rejecting traditional marital expectations to support herself through a software job. Similarly, in *Swagatam* [155], the male protagonist is a demanding customer at an arranged marriage matchmaking bureau. The manager of this bureau (coincidentally the female lead) has an online candidate repository, and tells the male lead to use the computer to filter through his requirements. She kindly reminds the hero that a woman's greatest trauma is being rejected at the arranged marriage meeting, and that technology should be used effectively to circumvent this problem, and thereby empower women. Given that marriage is practically an ever-present theme in Indian cinema, it is not surprising that several other films have used technology as a go-between for arranged or other marriages [151, 156].

5. CONCLUSION

"Distributing a film is like gambling without looking at your cards. We do not see the film, do not even read the script, only go by the actor's star pull, the director, music, and the production values. So the publicity matters a lot, since the first week is make-or-break for us. If it is a Vijaykanth film, we have posters with a family theme because you are attracting the rural audience. For an urban audience film, like Kandasamy, we have Vikram on the poster with a laptop."

Dr. Srinivasan, Abirami Film Distributors

The idea for this research emerged as after confounding outcomes in other research that indicated a mismatch between peoples' stated interest in technology and their actual use of state-provided computers in rural India. People were very excited about computers and the possibility of their own access to them, but unclear on how technology could be practically useful in their lives. Such ideas about technology were further seen to not just influence researchers' estimation of what the likely adoption for such projects may be, given the apparent enthusiasm about technology, but could also influence the populations' own propensities to invest in computer access projects, without a necessarily clear idea of the value of such technology. While there is no controlled intentionality in the use of images around technology for mass consumption, it was clear from discussions with filmmakers that they were deeply conscious of the importance of technology as an aspirational element in south Indian society.

Indeed many elements of the Indian cinema culture make it a unique case study drawing broader generalizations from it is consequently challenging.[6] However, it still stands true that the essential phenomenon discussed here, that of a mismatched enthusiasm about technology is prevalent in India just as much as in several other parts of the world. As subjects like community and social informatics spread their interest into the perception and consequent adoption of technology in developing regions, spreading our interests into analyses of media discourses is an inevitable direction of theoretical development.

6. REFERENCES

1. Higson, A., *The concept of national cinema*, in *Film and nationalism*, A. Wilson, Editor. 2002, Rutgers.
2. Foster, G., *Class-passing: social mobility in film and popular culture*. 2005: Southern Illinois Univ Pr.
3. Friedberg, A., *Window shopping: cinema and the postmodern*. 1994: Univ of California Pr.
4. Fehrenbach, H., *Cinema in democratizing Germany: reconstructing national identity after Hitler*. 1995: The University of North Carolina Press.

[6] Though Indian films are now consumed quite extensively throughout much of Latin America, Africa, and South East Asia.

5. Hopewell, J., *Out of the past: Spanish cinema after Franco*. 1986: British Film Inst.
6. Burton, J., *Cinema and social change in Latin America: conversations with filmmakers*. 1986: Univ of Texas Pr.
7. Lu, S., *Transnational Chinese Cinemas: Identity, Nationhood, Gender*. 1997: Univ of Hawaii Pr.
8. Meyer, B., *" Praise the Lord": Popular cinema and pentecostalite style in Ghana's new public sphere.* American Ethnologist, 2004. 31(1): p. 92-110.
9. McCall, J., *Madness, Money, and Movies: Watching a Nigerian Popular Video with the Guidance of a Native Doctor.* Africa Today, 2002. 49(3): p. 79-94.
10. Chakravarty, S., *National identity in Indian popular cinema, 1947-1987*. 1993: Univ of Texas Pr.
11. Gokulsing, K. and W. Dissanayake, *Indian popular cinema: a narrative of cultural change*. 2004: Trentham Books.
12. Rajadhyaksha, A., *The'Bollywoodization' of the Indian cinema: cultural nationalism in a global arena.* Inter-Asia cultural studies, 2003. 4(1): p. 25-39.
13. Boyd, D., *Why youth social network sites: The role of networked publics in teenage social life.* The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning, 2007: p. 119-142.
14. Nakamura, L., *Cybertypes: Race, ethnicity, and identity on the Internet*. 2002: Routledge.
15. Ignacio, E., *Building diaspora: Filipino community formation on the Internet*. 2005: Rutgers Univ Pr.
16. Wenjing, X., *Virtual space, real identity: Exploring cultural identity of Chinese Diaspora in virtual community.* Telematics and Informatics, 2005. 22(4): p. 395-404.
17. Bastian, M., *Nationalism in a virtual space: immigrant Nigerians on the internet.* West Africa Review, 1999. 1(1).
18. Fürsich, E. and M. Robins, *Africa. com: The self-representation of sub-Saharan nations on the World Wide Web.* Critical Studies in Media Communication, 2002. 19(2): p. 190-211.
19. Garrido, M. and A. Halavais, *Applying Social-Networks Analysis to Study Contemporary Social Movements.* Cyberactivism: Online activism in theory and practice, 2003: p. 165.
20. Burrell, J., *Problematic Empowerment: West African Internet Scams as Grassroots Media Production.* Information Technology and International Development, 2008. 4(4): p. 15-30.
21. Castells, M., *The Informational City: Information Technology, Economic Restructuring, and the Urban-regional Process*. 1991: Blackwell Publishers.
22. Castells, M., *Information technology, globalization and social development*. 1999, Geneva: UNRISD.
23. Roche, E.M. and M.J. Blaine, *Information technology, development and policy.* Information Technology, Development and Policy, EM Roche & MJ Blaine (eds.). Avebury: Aldershot, UK, 1996.
24. Mansell, R., *Knowledge Societies: Information Technology for Sustainable Development*. 1998: Oxford Univ Press.
25. Webster, F., *Theories of the Information Society*. 2002: Routledge.
26. Warschauer, M., *Technology And Social Inclusion: Rethinking the Digital Divide*. 2004: MIT Press.
27. Bailur, S., *Using Stakeholder Theory to Analyze Telecenter Projects.* Information Technologies and International Development, 2006. 3(3): p. 61-80.
28. Sein, M.K. and G. Harindranath, *Conceptualizing the ICT Artifact: Toward Understanding the Role of ICT in National Development.* The Information Society, 2004. 20(1): p. 15-24.
29. Franklin, M.I., *I Define My Own Identity: Pacific Articulations of'Race'and'Culture'on the Internet.* Ethnicities, 2003. 3(4): p. 465.
30. Rai, A.S., *India On-line: Electronic Bulletin Boards and the Construction of a Diasporic Hindu Identity.* Diaspora: A Journal of Transnational Studies, 1995. 4(1): p. 31-58.
31. Bastian, M.L., *Nationalism in a Virtual Space: Immigrant Nigerians on the Internet.* West Africa Review, 1999. 1(1).
32. Burrell, J. and K. Anderson, *"I have great desires to look beyond my world:" Trajectories of Information and Communication Technology use among Ghanaians.* New Media and Society, 2008. 10(2): p. 203-224.
33. Morton, H., *Islanders in Space: Tongans Online.* Small Worlds, Global Lives: Islands and Migration, 1999: p. 235–254.
34. Radhakrishnan, S., *Examining the "Global" Indian Middle Class: Gender and Culture in the Silicon Valley/Bangalore Circuit.* Journal of Intercultural Studies, 2008. 29(1): p. 7-20.
35. Kuriyan, R. and K.R. Kitner. *Constructing Class Boundaries: Gender and Shared Computing*. in *Second International Conference on Information Technologies and Development, Dec 2007, IEEE Conference Proceedings*. 2007. Bangalore.
36. Méliès, G., *Le Voyage dans la Lune*. 1902: France.
37. Chaplin, C., *Modern Times*. 1936, United Artists: USA.
38. Atkinson, P. *Technology and the Vision of Utopia*. in *Imagining the Future: Utopia, Dystopia and Science Fiction*. 2005. Monash University, Melbourne.
39. Bendle, M. *Zarathustra's Revenge: The Sordid Utopia of Contemporary Science Fiction Films*. in *Imagining the Future: Utopia, Dystopia and Science Fiction*. 2005. Monash University, Melbourne.
40. Youngblood, D.J., *Movies for the Masses: Popular Cinema and Soviet Society in the 1920s*. 1992: Cambridge University Press.
41. Singer, B., *Melodrama and Modernity: Early Sensational Cinema and Its Contexts*. 2001, New York: Columbia University Press.
42. Akudinobi, J., *Tradition/Modernity and the Discourse of African Cinema*. 1995, IRIS.
43. Lopez, A.M., *" Train of Shadows": Early Cinema and Modernity in Latin America.* Multiculturalism, Postcoloniality, and Transnational Media, 2003. 40(1): p. 48-78.
44. Mukherjee, R., *Hum Hindustani*. 1960.
45. Maté, R., *When Worlds Collide*. 1951, Paramount: USA.
46. Zaslove, M. and C. Nichols, *The Jetsons*. 1962, Hanna-Barbera Productions: USA.
47. Newman, S., *The Avengers*. 1961, American Broadcasting Company (ABC): USA.
48. Roddenberry, G., *Star Trek*. 1966, Paramount: USA.
49. Kubrick, S., *2001: A Space Odyssey*. 1968, MGM / Warner: USA.
50. Russel, K., *Billion Dollar Brain*. 1967, Jovera SA: UK.
51. Till, E., *Hot Millions*. 1968, MGM: UK.
52. Lang, W., *Desk Set*. 1957, 20th Century Fox: USA.
53. Bonnière, R., *Hide and Seek*. 1977, Public Broadcasting System (PBS): Canada.
54. Badham, J., *War Games*. 1983, MGM: USA.
55. Stone, O., *Wall Street*. 1987, 20th Century Fox: USA.
56. Faden, E., *The Cyberfilm: Hollywood and Computer Technology.* Strategies, 2001. 14(1).
57. Dwyer, R. and D. Patel, *Cinema India: The Visual Culture of Hindi Film*. 2002: Rutgers University Press.
58. Appadurai, A. and C. Breckenridge, *Public modernity in India.* Consuming Modernity: Public Culture in a South Asian World, 1995: p. 1–20.

59. Rajadhyaksha, A., *The Bollywoodization of the Indian cinema: cultural nationalism in a global arena.* Inter-Asia Cultural Studies, 2003. 4(1): p. 25-39.
60. Mankekar, P., *Screening Culture, Viewing Politics: An Ethnography of Television, Womanhood, and Nation in Postcolonial India*. 1999: Duke University Press.
61. Parvez, S., *Jumbish*. 1986, Ghalib Studio: India.
62. Srinivasa Rao, S., *Aditya 369*. 1991, Sridevi Arts: India.
63. Soni, S., *Shreeman Funtoosh*. 1965: India.
64. Ramachandran, M., *Ulagam Sutrum Valeban*. 1973: India.
65. Meher, S., *Babula*. 1985: India.
66. Roshan, R., *Koi Mil Gaya*. 2003: India.
67. Mahadevan, A., *Indradhanush*. 1989, Doordarshan: India.
68. Kapur, S., *Mr. India*. 1987: India.
69. Rajasekar, *Vikram*. 1986: India.
70. Kumar, R., *Chanakyan*. 1989: India.
71. Ratnam, M., *Roja*. 1992: India.
72. Singh, A., *Raqeeb*. 2007: India.
73. Ghosh, K., *Fida*. 2004: India.
74. Roshan, R., *Krrish*. 2006, Film Kraft: India.
75. Anand, S., *Bachna Ae Haseeno*. 2008, Yash Raj Films: India.
76. Ratnam, M., *Alai Payuthey*. 2000: India.
77. Roy, B., *Naukari*. 1954, Bimal Roy Productions: India.
78. Chandra, N., *Ankush*. 1986: India.
79. Bapaiah, K., *Aaj Ka Daur*. 1985: India.
80. Muthuraman, S., *Aarilirunthu Arubathu Varai*. 1979.
81. Kranthikumar, *Kanden Seethaiyai*. 2001: India.
82. Vasanth, *Satham Podathey*. 2007: India.
83. Renjith, *Nandanam*. 2002: India.
84. Narayana, K.L., *Rakhi*. 2006, Sri Durga Arts: India.
85. Selva, V., *Youth*. 2002, Ayngaran International: India.
86. Menon, R., *Kandukondain Kandukondain.* 2000.
87. Selvaraghavan, *Adavari Matalaku Ardhalu Verule*. 2007: India.
88. Bhargavan, *Thiru Ranga*. 2007: India.
89. Ravishankar, *Varushamellaam Vasantham*. 2002: India.
90. Menon, G., *Minnale*. 2001: India.
91. Rasool, *Unnai Paartha Naal Muthal*. 2004, Anandi Arts: India.
92. Yogi, *Chintakayala Ravi*. 2008: India.
93. Pandian, M.S.S., *The Image Trap: MG Ramachandran in Film and Politics*. 1992: Sage.
94. Jacob, P., *From co-star to deity: Popular representations of Jayalalitha Jayaram.* Women: A Cultural Review, 1997. 8(3): p. 327-337.
95. Sharma, A., *Blood, Sweat and Tears: Amitabh Bachchan, Urban Demi-God.* You Tarzan: Masculinity, Movies and Men, 1993: p. 167-180.
96. Shankar, S., *Sivaji*. 2007, AVM Productions: India.
97. Perarasu, *Dharmapuri*. 2006, A.M. Rathnam: India. p. 150 min.
98. Vidhyadharan, *Vaitheeswaran*. 2008, Annamalai Films: India.
99. Murugadoss, A.R., *Ramana*. 2002, Oscar Films: India.
100. Vinayak, V., *Tagore*. 2003, Oscar Movies: India.
101. Vasu, P., *Paramasivan*. 2006: India.
102. Vishnuvardhan, *Billa*. 2007, Sri Keerthi Creations: India.
103. Shankar, S., *Anniyan*. 2005, Lakshmi Ganapathi Films: India.
104. Jayaraj, R., *4 The People*. 2004: India.
105. Reddy, G., *New populism and liberalisation: regime shift under Chandrababu Naidu in AP.* Economic and Political Weekly, 2002. 37(9): p. 871-883.
106. Rudolph, L. and S. Rudolph, *The Iconization of Chandrababu: Sharing Sovereignty in India's Federal Market Economy.* Economic and Political Weekly, 2001. XXXVI(18): p. 1541-1551.
107. Pani, N., *Icons and Reform Politics in India: The Case of SM Krishna.* Asian Survey, 2006. 46(2): p. 238-256.
108. Menon, R., *Kandukondain Kandukondain*. 2000: India.
109. Pande, R., *Looking at information technology from a gender perspective: the call centers in India.* AJWS, 2005. 11(1): p. 58-82.
110. D'Mello, M., *Gendered selves and identities of information technology professionals in global software organizations in India.* Information Technology for Development, 2006. 12(2): p. 131-158.
111. Patel, R. and M. Parmentier, *The Persistence of Traditional Gender Roles in the Information Technology Sector: A Study of Female Engineers in India.* Information Technologies and International Development, 2005. 2(3): p. 46.
112. Khan, M., *Mother India*. 1957: India.
113. Roy, B., *Madhumati*. 1958, Bimal Roy Productions: India.
114. Ramanna, *Jawab*. 1970: India.
115. Bhimsingh, A., *Sila Nerangalil Sila Manithargal* 1975: India.
116. Bedi, N., *Adalat*. 1976: India.
117. Chopra, Y., *Deewar*. 1975: India.
118. Krishnamurthy, R., *Thee*. 1981: India.
119. Majumdar, P., *Aarti*. 1962, Rajshri Productions: India.
120. Sahu, K., *Dil Apna Aur Preet Parai*. 1960, Mohan Pictures: India.
121. Balachander, K., *Manathil Uruthi Vendum*. 1987: India.
122. Chopra, Y., *Daag*. 1973: India.
123. Padmarajan, P., *Koodevide*. 1983: India.
124. Kapoor, R., *Shree 420*. 1955, RK Films: India.
125. Kapdi, M., *Aaitya Bilavar Nagoba*. 1979: India.
126. Prakash Rao, K.S., *Secretary*. 1976: India.
127. Balachander, K., *Avargal*. 1977: India.
128. Chopra, B.R., *Pati Patni Aur Woh*. 1978, B.R. Films: India.
129. Rawail, H.S., *Patanga*. 1949, Varma Films: India.
130. Neelakantan, P., *Raman Thediya Seethai*. 1972, AVM Productions.
131. Paranjpe, S., *Chashme Buddoor*. 1981, PLA Productions: India.
132. Kanwar, R., *Laadla*. 1994, Neha Arts: India.
133. Mirza, A., *Raju Ban Gaya Gentleman*. 1992, Sippy Films: India.
134. Rama Rao, T., *Andhaa Kanoon*. 1983: India.
135. Dixit, C.P., *Fakira*. 1976, Eros Entertainment: India.
136. Bhogal, A., *Zakhmi Aurat*. 1988, Manta Movies: India.
137. Chopra, B.R., *Insaf Ka Tarazu*. 1980, B.R. Films: India.
138. Sippy, R., *Shakti*. 1983: India.
139. Hussain, N., *Yaadon Ki Baaraat*. 1973, United Producers: India.
140. Balachander, K., *Arangetram*. 1973: India.
141. Sasi, I.V., *Avalude Ravakul*. 1980: India.
142. Samanta, S., *Amar Prem*. 1971, Shakti Films: India.
143. Ramachandran, M.G., *Ulagam Sutrum Valiban*. 1973, Emgeeeaar Pictures: India.
144. Anand, V., *Jewel Thief*. 1967, Navketan Films: India.
145. Bhadran, *Olympian Anthony Adam*. 1999: India.
146. Madan, *Pellaina Kothalo*. 2004, AA Naluguru Films: India.
147. Selvaraghavan, K., *Aadavari Matalaku Ardhalu Verule*. 2007, Sri Saideva Productions: India.
148. Yelleti, H., *Ankit, Pallavi and Friends*. 2008: India.
149. Goud, R.K., *Jodi No. 1*. 2003, RK Movies: India.
150. Shankar, S., *Kunjikkoonan*. 2002: India.
151. Prabhakar, T., *Neethone Vuntanu*. 2002, Lakshmi Art Pictures: India.
152. Vamsy, *Avunu Valliddaru Ista Paddaru*. 2002, Anandi Arts: India.
153. Selvaraghavan, *7G Rainbow Colony*. 2005, Ayngaran: India.
154. Kammula, S., *Anand*. 2004: India.
155. Dasarath, *Swagatam*. 2008, Aditya Ram Movies: India.
156. Kathir, *Kadhalar Dhinam*. 1999, A.M. Rathnam: India.

Towards Trusted Cloud Computing

Jerry Robinson and Joon S. Park
Syracuse University
342 Hinds Hall
Syracuse, NY 13244

{jlrobi02, jspark}@syr.edu

ABSTRACT

In this paper we analyze the current cloud-computing architectures and discuss their challenges for more trusted services in the future. First, we compare cloud computing to related technologies, including distributed computing, grid computing, and Web services. Then, we analyze the current architectures of cloud computing with examples of how it is currently utilized and identify a new composite architecture. Finally, we discuss the technical, security, policy, and other issues towards trusted cloud computing for its widespread adoption and use.

Categories and Subject Descriptors

*C.2.4 [**Computer-Communication Networks**]: Distributed Systems—distributed applications; D.4.6 [**Operating Systems**]: Security and Protection—information flow controls*

General Terms

Reliability, Performance, Security

Keywords

Cloud computing, infrastructure models, security, trust

1. INTRODUCTION

Cloud computing is a relatively new and blossoming computing platform. It offers scalable IT services over the Internet in the form of infrastructure, software developer platforms, and hosted Web applications. The primary appeal for organizations moving their data and applications to the cloud is the freedom from having to devote significant monetary investment, human resources, and other assets to building and maintaining a data center, multiple developer environments, and in-house Web application hosting and scaling. [7].

Cloud computing refers to the delivery of applications as services over the Internet as well as actual cloud infrastructure (the hardware and systems software in the virtual data centers that provide these services) [2]. Cloud computing allows users to build, access, and utilize applications as well as store and access data through the Web [8]. In a cloud computing arrangement, computing and/or data reside in virtual data centers operated by cloud providers. Cloud computing gives the illusion of infinite and scalable computing resources that are available on demand.

The basic premise of cloud computing is concentrated computation and storage in a carefully managed core, where high-bandwidth connections link high performance machines [9]. As with most computing platforms end users both make requests that initiate computations and receive the results of those computations [9].

Cloud computing as it exists today is the result of several technological developments including reductions in the cost of storage and client CPU bandwidth, HTML, CSS, AJAX, and REST, virtualization, utility computing, and Service Oriented Architectures [7]. The three primary predecessor technologies of cloud computing are distributed computing, grid computing, and Web services.

2. COMPARISON WITH RELATED WORK

In this section, we compare cloud computing to related technologies, including distributed computing, grid computing, and Web services.

2.1 Distributed Computing

The main distinction between grid and distributed computing is the way resources are managed. Distributed computing uses a centralized resource manager and all nodes cooperatively work together as a single unified resource or a system [11, 13]. In grid computing, each node has its own resource manager and the system does not act as a single unit. While grid computing usually lightens the load on heavily utilized machines, distributed computing is typically utilized in order to process or run complex and resource draining programs quickly and efficiently. Also, while grid computing enables a computer to run its applications on other networked machines, computers in most distributed environments communicate by message passing and have independent memories and operating systems [11].

2.2 Grid Computing

We summarize the primary distinctions between cloud computing and grid computing as follows [3, 5, 20].

- Grids are made up of networked high-end computer servers/clusters. Clouds are composed of commodity computers, high-end servers, and network attached storage.
- Grid nodes are interconnected via the Internet with high latency and low bandwidth. Cloud nodes are interconnected via high-end dedicated network with low latency and high bandwidth.
- Grids are open, meaning it is easy for a node to switch grids. Clouds, on the other hand, are not open. Apps built on the cloud infrastructure and data stored in the cloud cannot be easily moved to the cloud of a different provider.
- Grid applications are accessed via grid middleware while cloud applications are accessed via the Internet standard web protocols.
- Grids are often used for scientific collaboration and data or compute intensive applications. Cloud computing can also provide these abilities in addition to providing scalable infrastructure, platform, and software services over the Internet.

- Grids are restricted to the software applications available within the grid environment. Cloud computing allows any network connected computer with a web browser to access and utilize any application residing in the cloud that a user have access to.

2.3 Web Services

The third predecessor technology to cloud computing is Web services. Web services convert standard applications into web applications that are publishable, found, and used on the World Wide Web via any browser on any platform. Web services use XML to code and decode data so that it can be exchanged between different applications and platforms. They also allow a developer to create reusable application components that can be integrated into a web application (w3schools.com, 2009).

The basic Web service platform is XML + HTTP. This platform enables application interoperability by providing a way for applications to exchange data. XML serves as the go between language for different platforms and programming languages (w3schools.com, 2009). Simple Object Access Protocol (SOAP) is the transport vehicle for this data. SOAP is an XML based protocol that lets applications exchange messages over HTTP (w3schools.com, 2009).

Web Services Description Language (WSDL) is an XML based language for describing web services and how to access them. The WSDL document defines a web service, its performable operations, the messages and data types used by the web service, and the communication protocols used (w3schools.com, 2009).

3. ANALYSES ON CLOUD COMPUTING ARCHITECTURES

Clouds have three formal architectural layers. The lowest layer is infrastructure and it delivers basic storage and compute capabilities over a network. The middle platform layer includes an integrated environment which provides abstractions and services to develop, test, deploy, host, and maintain applications. The application layer is the highest layer where a complete application is offered as a service over the Internet [8]. There are also two informal cloud architectural layers. An organization can utilize all three formal layers to build applications on top of configurable infrastructure and have those applications deployed from the cloud over the Internet. We call this the composite service model. Alternatively, organizations can combine one or more cloud architectures with their own in-house resources. This is known as the hybrid model.

3.1 Infrastructure-as-a-Service and Infrastructure Models

Infrastructure as a Service (IaaS)– Resources such as servers, storage systems, switches, routers, computing instances, databases, networks, load balancers, etc. are pooled and offered as a service to handle varying intensities of workloads [6]. These infrastructure resources are rented by the consumer who either can use them as needed or can build, deploy, and run applications (including operating systems and applications) on top of them. Examples include Amazon Web Services' Elastic Compute Cloud (E2C) which provides compute and storage infrastructure.

IaaS offerings eliminate the need for organizations to build and maintain their own data centers. Rather than making large capital investment to own and operate a data center that is likely to be underutilized during periods of minimal activity or over utilized during peak activity periods, IaaS allows the user to pay for infrastructure resources based on actual usage. Cloud providers own and manage infrastructure resources and offer them as an integrated, pay-as-you-go service [4]. The resources can be accessed via APIs which allow IaaS customers to configure them as needed [4].

Customers can access and configure infrastructure services via software APIs. Virtualization technologies at the IaaS level allow users to run specified virtual machine instances. For example, a customer can run either the Windows Operating System (OS) instance or the Solaris OS [4].

Infrastructure Models

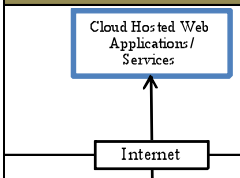
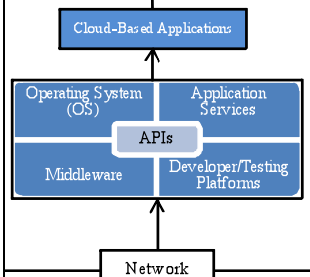
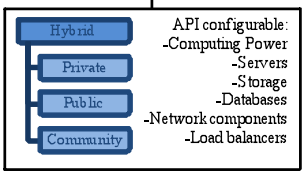
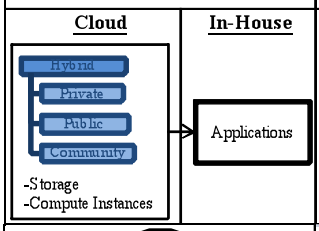
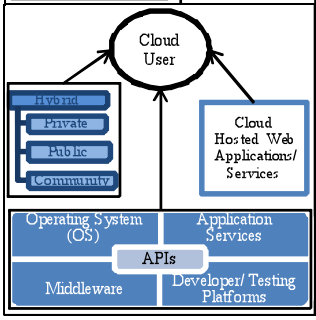
At the IaaS layer users subscribe to one of four infrastructure models. Each model offers a different level of access, security, and ownership with regard to the infrastructure resources (servers, storage systems, switches, routers, etc.) and stored data. The choice impacts the level of security, privacy, scalability, and QoS that a customer should expect.

Public cloud – Cloud infrastructure is owned by the cloud provider who rents infrastructure services to the general public [15]. Applications from different customers may be mixed together on the cloud's servers, storage systems, and networks [6]. Similarly, all data placed in a public cloud is stored together regardless of where it originates. In other words, data belonging to different users is stored in the same place. Public clouds are often much larger than private or community ones and offer the ability to scale up or down on demand.

*Note: Portions of public clouds can be carved out and utilized exclusively by one customer, creating a virtual private datacenter.

The components of the virtual private data center can be located in the same facility as the customer giving an organization greater control over its data and access to its own bandwidth resources within the data center. Virtual private datacenters give customers greater visibility into cloud infrastructure. In addition to virtual machine images developers can manage servers, storage systems, network devices, and network topology. [6].

Private cloud – Cloud built for the exclusive use of one customer to provide greater control over data, security, and quality of service. The organization/user owns the infrastructure and controls how applications are deployed on it [6]. Private clouds can be managed in the organization's datacenter by its own IT organization or in the cloud provider's facility [6].

Diagram	Architectural Layer	Description	Examples	Key Characteristics
	Software-as-a-Service (SaaS)	Hosted Web applications and services accessible by end users through a web browser or web service interface.	-GoogleDocs -Salesforce.com sales and CRM applications -Google's Gmail, Yahoo! Email service	-Eliminates need to install and run applications on the end-user's computer. -Eliminates need for upfront software purchases
	Platform-as-a-Service (PaaS)	An integrated solution that allows developers to build and test applications and services from the cloud.	-Google AppEngine -Microsoft Azure	-Eliminates need for separate application build, test, and delivery environments. -Applications are created using programming languages and tools (e.g., java, .net, python, etc.) -Ability to create mashups
	Infrastructure-as-a-Service (IaaS)	Computing infrastructure offered as a service. Cloud infrastructure models include: 1) Public 2) Private 3) Community 4) Hybrid.	-Amazon Elastic Compute Cloud (EC2)	-Eliminates need for company owned data centers. -Accessible and configurable infrastructure services via software APIs. -Infrastructure resources utilized as needed in the form of compute instances and/or storage
	Hybrid Service	The combined utilization of one or more cloud services with in-house resources.	-Computing instances purchased to support in-house applications during peak times -Developer deploys in-house applications that send/receive data from IaaS storage	-Enables gradual adoption of cloud computing. -Ideal for organizations needing infrastructure or other resources on an irregular basis
	Composite Service	The combined utilization of IaaS, PaaS, and Saas. It entails using the cloud: 1. for compute and storage instances, 2. to build, test, and deploy applications, 3. to host developed applications as web applications and 4. to deliver these web applications across the internet	-Computing infrastructure, a developer's platform, and web hosted applications all offered as a service.	-Eliminates need for company owned data centers, -Eliminates need for separate application build, test, and delivery environments. -Eliminates need for application installation and upgrades on end-user PCs.

Community cloud – Cloud built for the exclusive use of a specific group or community of users. The group can be composed of several organizations or units within an organization that share the same or similar goals, missions, security requirements, policy, and compliance requirements [15].

Hybrid cloud – A combination of two out of the three previously mentioned models. Hybrid clouds help to ensure the right scalability/security mix. For instance, augmenting a private cloud with the resources from a public one allows an organization to maintain service levels when workflow fluctuates while retaining control over data. The most common scenario for the hybrid model is when storage clouds are used to support web 2.0 applications [6].

3.2 Platform as a Service

Platform as a Service (PaaS) – Includes a layer of software offered as a service that can be used to build higher-level applications and services [6]. It is an integrated solution that allows developers to build and test applications and services on the same platform within the cloud. Commercial PaaS examples include Google Apps Engine and Microsoft Azure. AppEngine allows users to build software applications on top of Google's

infrastructure. Azure offers SQL, .Net, and windows based services to application developers.

Applications can be created using programming languages and tools (e.g., java, .net, python, etc.) that the cloud provider supports [15]. Alternatively, PaaS providers offer visual customizations that developers can use to create a software application instead of writing code [15]. Developers are also able to create mashups by integrating common web services (Google Maps, Google Calendars, weather services, etc.) with their applications. The PaaS provider might integrate an OS, middleware, application software, and a development environment that is then offered to a customer as a service. Developers see the cloud provider's service presented to them through an application programming interface (API) or graphical user interface (GUI) [6].

One of the critical components at the PaaS levels is the Application Programming Interfaces (APIs). APIs allow developers to control how the cloud infrastructure supporting PaaS applications is utilized. Developers can specify how virtual components (servers, storage, and network resources) are configured and interconnected, and how virtual machine images and application data are stored and retrieved from storage clouds. APIs also allow developers to configure the platform service offering so that it does what it necessary to scale itself in order to provide a chosen level of service.

3.3 Software as a Service

Software as a Service (Saas) - Complete cloud applications hosted as web applications or services and made available to end users over the Internet [4]. The software runs on the cloud (not on user PCs) and multiple end users can utilize it at the same time [6]. The applications are accessible from any web-connected device [15]. The application can be hosted on any Internet exposed data center. They are commonly hosted by either a SaaS software vendor on their web servers or an infrastructure service provider. Examples include salesforce.com and Google Docs, Salesforce.com sales and CRM applications, Google's Gmail, Yahoo! Email service, and billing and monitoring services

SaaS offerings eliminate the need to install and run applications on the end-user's computer. Instead, the end user can access and use applications residing in the cloud via an Internet browser or web service interface. SaaS also eliminates the need for upfront software purchases by the end user. SaaS providers allow customers to take advantage of on-demand pay-as-you-go application hosting services. Maintenance and updates are handled by the cloud provider. When software is upgraded the SaaS provider makes sure that new software version is compatible with existing data [4].

Hybrid Service – The combined utilization of one or more of the three previously mentioned architectural layers with in-house resources. This model allows organizations to move to the cloud in phases. It allows the user to choose the cloud services that are most needed at a given time in a given situation. An example would be if a customer built its own in-house applications with IaaS compute instances and infrastructure supporting them [18]. Another example would be the utilization of computing instances both to handle process intensive imaging and to store them in one of the organization's queues.

Hybrid Service offerings are particularly useful for organizations that are not completely ready or willing to move all of their data and/or applications to the cloud. Hybrid Service enables gradual adoption of cloud computing. It is also ideal for organizations that need infrastructure or other resources on an irregular basis and do not want to manage an entire data center that is likely to sit idle most of the time.

Composite Service (CS) - The combined utilization of IaaS, PaaS, and SaaS by one customer. Users of the CS model utilize the cloud to exploit infrastructure resources, build, test, and deploy cloud-based applications on top of those resources, and have their applications hosted as web applications and services.

CS eliminates the need for company owned data centers, separate developer environments, and application installation and upgrades on end-user PCs. This combined offering allows cloud users to capitalize on all of the benefits that each cloud computing service offers.

While CS is not formally recognized as an architectural model, it is likely that cloud providers will offer this integrated solution as more organizations move to the cloud. Composite service subscribers are able to store data, configure infrastructure components, build higher level scalable applications on top of the platform covering the infrastructure, and have the ability to add, change, and delete their cloud applications hosted on the web.

4. IMPACTS AND CHALLENGES

4.1 Impacts

One of the greatest potential advantages of cloud computing is that it allows organizations to avoid devoting significant monetary investment, human resources, and other assets into building and maintaining needed application services and the required infrastructure [7]. Clouds enable IT resources to be configured according to a user's needs [19]. Customers have the flexibility to make changes to their architectures in order to scale service when needed without making huge capital expenditures [7]. Developers using a cloud are able to develop, test, and run software on any computing platform offered by the provider [9]. The cloud model allows customers the option to pay a monthly subscription fee or usage rate rather than pay the high cost of setting up and maintaining a technology infrastructure. Software residing in a cloud can be upgraded easily and quickly by the cloud provider. As long as there are no service interruptions SaaS users are able to access applications through any web browser or Internet connected device at any time.

4.2 Challenges

4.2.1 Technical

Before widespread adoption of cloud computing can occur, several technical obstacles must be overcome. The first obstacle is the lack of standards for cloud provider APIs. Customers cannot easily move their data and applications from one provider to another because doing so requires them to learn both a new set of APIs and new methods for configuring infrastructure components to support the optimal performance of their applications [2]. While cloud providers are likely to prefer the ability to lock customers into their offerings, most cloud users are likely to prefer the ability to deploy their applications on any provider's infrastructure quickly and easily in case of service

price increases, reliability problems, or the extinction of their provider [2].

The second obstacle is the reliability of service. Organizations rely heavily upon their technical infrastructures for continued existence. A temporary service outage can have devastating effects on an IT dependent company utilizing the cloud for infrastructure, platform, or software services. In a February 2009 UC Berkley research paper outage data for 2008 was presented for Amazon Simple Storage Solutions (S3), Google AppEngine, and Gmail. The duration of the outages ranged from 1.5 to 8 hours *[2]*. The lack of interoperability between different cloud providers creates a burden for the customer if the cloud provider's services become unreliable. The customer does not have any guaranteed protections against losses incurred from missed business opportunities. Also, if service reliability is not acceptable to the customer time and effort must be invested into finding a different provider, moving data and/or applications to that provider, and learning a different set of APIs.

Third, unlike corporate data centers where security threats can have devastating impacts, cloud providers are able to mitigate the effects of these threats by replicating customer data and applications across multiple data centers. However, the threat of security breaches is not totally eliminated in the cloud environment. Since cloud providers host data and applications from several different customers, the impact of a security breach can lead to compromised data for several different organizations. In addition, while a failure in one data center will not necessarily lead to a total shutdown of service, it could lead to significantly reduced application performance [4].

Service-level-agreements govern cloud services. Unfortunately, SLAs do not dynamically adjust to the needs of the users (such as if a user needs to use more bandwidth than the SLA allows). Deploying an autonomous system to efficiently provision services in a cloud infrastructure is a challenging problem due to the unpredictability of consumer demand, software and hardware failures, dissimilarity of services, power management, and conflicting signed SLAs between consumers and providers [8].

The data transfer latency times between data centers in the cloud and clients accessing that data are higher than when data and applications reside in corporate data centers [4]. Data transfer bottlenecks are more common for SaaS cloud offerings (combined with the use of cloud storage facilities) and thus require cloud users and providers to consider the proximity of the user to the data center as well as the size of the spectrum pipe between the user and the data center [4]. A short and wide pipe will minimize the distance between the data source and destination and a wider pipe will allow more data packets to flow at a given time.

4.2.2 Security

Security, privacy, and reliability are major concerns with the movement towards cloud computing. Third-party possession of personal documents raises questions about control and ownership such as the transferability of content should the user switch providers, the loss of content for non-payment, and the ability to destroy unwanted documents [9]. Currently, there are no standards governing the cloud provider industry with regard to security, privacy, and reliability. As a result, each provider has full discretion over how they will manage these important issues [10].

Laws and regulations such as the USA Patriot Act require cloud providers to release personal and other user data to government authorities in the case of a search warrant/subpoena. Data residing in a public cloud is not secured or maintained by the owner which means that he/she has no say or means of recourse when governmental or law enforcement agencies (whether in the owner's country or somewhere else) demand delivery of their data. Additionally, some governments outside of the U.S. do not want data generated within their borders to reside in the U.S. because of the Patriot Act. This places a data portability constraint on cloud computing. If a provider does not have a data center in a particular country where a potential customer has operations it may not make sense for the potential customer to utilize that provider's services [4].

A major security obstacle to widespread adoption of cloud computing is the lack of provider tools that guarantee the confidentiality and auditability of data moved to and stored in the cloud. The majority of cloud offerings are public and multitenant, which means different users are accessing the same services and using the same infrastructure without the applications knowing about each other [4]. Even if the cloud has strong security measures in place, the ultimate security of all data is only as strong as the weakest link among the users. One user's security vulnerability can become the entire cloud's headache. Centralized services open the door for security threats in resource provisioning or during distributed application execution. Ironically, cloud computing can be used to carry out Internet crime as well. Hackers can use virtualized infrastructures as a launching pad for attacks [8].

Currently, cloud providers do not have any technical solutions that can guarantee their cloud will not become a victim to harmful malware, virus infection, hackers, or distributed denial of service attacks. Cloud providers with multiple data centers that replicate user data and applications may be able to limit the operational impact of harmful attacks [4]. It is unlikely that a customer's data will be lost or unavailable to the customer if an attack occurs. However, multiple data centers and replication do not protect against service outages or delays that may result. Also, multiple data centers and replication does not eliminate the possibility of data in any one center being compromised during an attack. Some cloud providers such as Sun Microsystems encourage their customers to encrypt data transmitted to and/or stored in the cloud to ensure that if data falls into the hands of a) intruders who are able to break through the cloud's security, b) unauthorized parties who are able to access user data because of a configuration error, or c) unauthorized parties able to gain access as the information passes over the Internet it cannot be interpreted [6].

4.2.3 Policy

There are a wide range of policy issues related to the cloud operating model that have not yet been resolved. User likely will neither want their content or information to be monitored or used by cloud provider or third parties [10]. However, there are no current policy guidelines protecting customers from unauthorized sharing or monitoring of personal information about the users or their usage statistics.

The global dispersion of cloud data centers creates the need to account for differing sets of regulations and privacy concerns in different parts of the world [10]. The differences in surveillance and copyright laws illustrate this concept. Surveillance is difficult

in the cloud environment because data centers are often located in countries other than the data owners and these countries often have conflicting laws regarding law enforcement's ability to search and seize personal property. Similarly, government enforcement of copyright laws could force cloud providers to take measures to ensure that customers do not illegally share copyrighted materials, which could reduce the practicality of offering cloud services to the general public [17].

Currently, licensed software can only installed and utilized on a certain number of computers [2]. It has not yet been determined whether redistribution includes using a particular licensed software product at work/home/school/etc. as well as in the cloud [10]. The current software licensing scheme hinders the use of licensed software in the cloud since it may be spread over hundreds or even thousands of compute instances rather than one or a few PCs.

Telecommunication policy neither accounts for content lost in transmission nor provides a framework for categorizing clouds. There is much debate on whether cloud providers should be considered ISPs, telecommunications providers, or common carriers. In addition, the network neutrality debate may impact the growth of cloud computing. If network neutrality is not guaranteed, the telecommunications service providers that control the underlying network connections would have the ability to limit a cloud provider's service through pricing and distribution structures, which could reduce the profits of cloud providers [10].

Currently, several other privacy and security related questions concerning cloud computing remain unanswered and policy guidance is still forthcoming. Examples include: when does information gathering cross the line and become unethical data mining? If a cloud user participates in illegal, harmful, or publicly undesirable activity (i.e., sending spam or executing DDOS attacks) does liability remain with the customer or transfer to the provider? In the U.S., where there is no universal standard for privacy protection, no determination has been made as to what type of privacy is guaranteed in the cloud computing, if any, or if it will be up to the individual providers to decide.

5. CONCLUSION AND FUTURE

In this paper we have analyzed the current cloud-computing architectures and discussed their challenges for more trusted services in the future. We have compared cloud computing to related technologies, including distributed computing, grid computing, and Web services. We have also analyzed the current architectures of cloud computing with examples and identified a new composite architecture. Finally, we have discussed the technical, security, policy, and other issues towards trusted cloud computing for its widespread adoption and use.

Cloud computing represents a shift in the way organizations manage their data and applications, developers design and deploy applications, and users access, transmit, and share information. Although the cloud model has yet to reach its full potential to deliver scalable services over the Web to many users, its continued and wide-spread growth could be forever hindered by the challenges described in this paper. The findings presented should serve as a foundation for further research that will uncover solutions to the problems identified and lead to increased usage of the cloud computing model.

6. REFERENCES

[1] (n.d.). Amazon Web Services. Retrieved from http://aws.amazon.com

[2] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., et al. (2009). Above the clouds: A berkeley view of cloud computing. EECS Department, University of California, Berkeley, Tech.Rep.UCB/EECS-2009-28.

[3] Bertis, V. (2002). Fundamentals of grid computing. IBM Redbook Paper. Austin, TX, November 11, 2002. http://www.redbooks.ibm.com/redpapers/pdfs/redp3613.pdf

[4] Bhattacharjee, R., & Cusumano, M. (2009). Massachusetts Institute Of Technology). An Analysis of the Cloud Computing Platform.

[5] Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems, 25(6), 599-616.

[6] Carolan, J., & Gaede , S., et al. (2009). Introduction to cloud computing architecture Sun Microsystems white paper.

[7] Computing, D., & Creeger, M. (2009). Cloud computing: An overview. Distributed Computing, 7(5)

[8] Dikaiakos, M., Katsaros, D., Mehra, P., Pallis, G., & Vakali, A. (2009). Cloud Computing - guest editors' introduction. IEEE Internet Computing, 13(5), 10.

[9] Hayes, B. (2008). Cloud computing. Communications of the ACM, 51(7), 9.

[10] Grimes, J. M., Jaeger, P., & Lin, J. (2008). Cloud computing and information policy: Computing in a policy cloud? Journal of Information Technology Politics, 5(3), 269.

[11] Godfrey, Bill (2002). A primer on distributed computing. Retrieved from http://www.bacchae.co.uk/docs/dist.html

[12] (n.d.). Google AppEngine. Retrieved from http://code.google.com/appengine/

[13] Kshemkalyani, A. D., & Singhal, M. (2008). Distributed Computing :Principles, Algorithms, and Systems. Cambridge: Cambridge University Press.

[14] Mell, P., & Grance, T. (October 7, 2009). Effectively and Securely Using the Cloud Computing Paradigm. Retrieved from http://www.google.com/url?sa=t&source=web&ct=res&cd=7&ved=0CB4QFjAG&url=http%3A%2F%2Fcsrc.nist.gov%2Fgroups%2FSNS%2Fcloud-computing%2Fcloud-computing-v26.ppt&rct=j&q=Effectively+and+Securely+Using+the+Cloud+Computing+Paradigm&ei=p6cAS47zINW3lAem-_WVCw&usg=AFQjCNG9p1P3Y5lDxEwh1WG1k7cdEGW9FA

[15] Mell, P., & Grance, T. (May 9, 2009). NIST Working Definition of Cloud Computing. Retrieved from http://groups.google.com/group/cloudforum/web/nist-working-definition-of-cloud-computing

[16] (n.d.). Microsoft Azure. Retrieved from http://www.microsoft.com/windowsazure/

[17] Nelson, M. (2009). The cloud, the crowd, and public policy. Issues in Science and Technology, 25(4), 71.

[18] O'Neil, V. (2009). Connecting to the cloud, part I: Leverage the cloud in applications. Take advantage of the hybrid model. *IBM Developer Works* Retrieved from http://www.ibm.com/developerworks/library/x-cloudpt1/

[19] Oppenheim, R. (2009). A match made in heaven: Cloud computing and mobile technologies. Searcher, 17(7), 14.

[20] Weinhardt, C., Anandasivam, A., Blau, B., & Stößer, J. (2009). Business models in the service world. IT Professional, 11(2), 28-33.

Empirically Assessing Impact of Scholarly Research

Jian Qin
School of Information Studies
Syracuse University
Syracuse, NY
315-443-5642
jqin@syr.edu

ABSTRACT

The impact of scholarly research can be manifested in many different ways, but conventional citation measures are often used to evaluate research impact without distinction. This paper attempts to analyze the misconception about what research impact is and what exactly citation data measures or cannot measure for impact. The author proposes a framework of impact assessment, in which research output makes intellectual, technological, and societal impact through knowledge diffusion. The extent of knowledge diffusion, adoption of technology and practices, and benefits from adoption constitute the overall impact of research. Although still in its preliminary form, the framework offers a more holistic view on the composite of research impact.

General Terms

Measurement

Keywords

Impact assessment, citation-based measures, scholarly research

1. INTRODUCTION

Citation data is empirical in nature and has been used to study the extent to which a paper or a journal was cited. It has been used to quantify a wide range of things, ranging from the evaluation of research quality and impact to the mapping of science. A large number of publications have been produced on all these topics since the *Science Citation Index* started half a century ago. While using citations to map science is considered a useful approach and can provide valuable "big picture" [11], the debate on the validity and reliability of citations as measures for impact assessment never reached an agreement. Concerns on citation measures for research performance and impact primarily come from the inherent limitations of citation database: inadequate or biased coverage for countries, disciplines, and languages of research publications [1] [9], as well as the ambiguities and confusions caused by name abbreviations and orders of author names. Such concerns and critiques, however, rarely question the meaning of impact: What does impact of scholarly research mean exactly? How does citation data measure the impact?

The word "impact" has been used loosely to refer several things. We frequently read in literature that the number of citations a paper received is considered as "research impact," thus the average number of citations received by a research group would be "average impact" [10], or "indices of scientific impact" [5]. These impact measures, however, are vague on what they mean exactly and lack elaboration on theoretical implications. While citation counts reflect the extent to which a research publication is known or visible to the research community, the data does not tell what role a cited work played in the creation of citing work, nor does it show whether it received criticisms or served as the "giant's shoulder" for the citing work. Citation counts may be a "proxy for the objective quality of an article" (Oswald, 2009), but far from telling the whole story of research impact.

What does research impact mean? How can it be measured? This paper attempts to analyze the composite of research impact with a focus on information science research. Unless specified, all discussion in this paper is placed in the context of information science research. In the framework proposed in this paper, three factors are discussed—extent, adoption, and benefits—along the intellectual, technological, and societal aspects of research impact.

2. THE COMPOSITE OF RESEARCH IMPACT

Impact implicates a strong influence, effect, or a forceful consequence. Impact from natural environment, such as that of invasive species on ecosystem, or hurricanes on affected regions, is concrete and visible, hence easier to measure in the economic, ecosystem, and health terms. Impact of scholarly research (and especially in information science), on the contrary, is not always easily measurable by economic gains or losses, or in countable figures. In addition, research impact requires some ingredients to brew: the utilization of research output, be it papers, data, patents, software, or otherwise, and a process of diffusion of these research outputs through humans and social and economic activities (Figure 1).

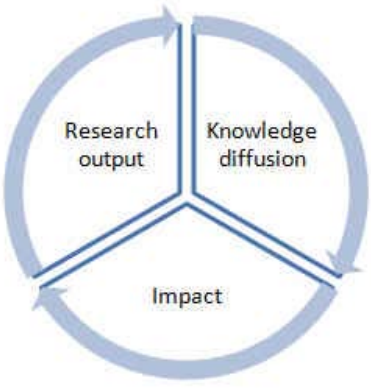


Figure 1. The research impact cycle

The boundaries between research output and knowledge diffusion are often not clear-cut and different models of impact assessment have been applied in various contexts. The impact assessment case studies at the Arts and Humanities Research Council (AHRC) (UK) apply a logic model of assumptions, resources, processes/activities, outputs, outcomes, and impact to evaluate the impact of their funded research projects [8]. Duryea, Hochmanand and Parfitt [4] define a similar model in which they specify research outputs, research transfer, research outcomes, and research impact. These models provide supportive footnotes to Figure 1 that knowledge diffusion does not start only when research output is produced; rather, it may well begin when a research idea is being formulated and the interaction between output from previous research and ideas/theories from current research lays the foundation for diffusion of knowledge and impact.

The orientation of research has a direct effect on what types of impact will result from research. Theory-based hypothesis-testing research, for example, would produce primarily intellectual impact, which may or may not trigger other types of impact later along the stages defined by the diffusion theory's adoption process—knowledge, persuasion, decision, trial, and adoption [7]. Another case is the practice-based research. Research with this orientation in information science typically involves developing algorithms, software, databases, standards, best-practice guidelines, etc. that can be applied or implemented to benefit work processes and/or customers. When output from practice-based research is diffused, the measures for research impact include productivity increased, time saved, or revenues gained/cost saved [4]. Even though these aspects of impact are relatively easy to quantify, the data is not always easy to collect. This is particularly true for information science research because the research output may be used by any sectors and the effects or benefits of using them are recorded elsewhere other than the institution where such output is produced.

The field of research can also determine how research output might make an impact. The Backer Medical Library at Washington University in St. Louis developed a model for faculty to assess the impact of their research, in which the impact of their research can be measured by community benefits from the diffusion and adoption of their research output. Such community benefits include 1) economic outcomes indicated by a cost-effective intervention for a disease, condition or disorder among other things, 2) health care outcomes as reflected in clinically effective approach in the management and treatment of a disease, disorder or condition, and 3) enhancement of quality of life [2]. In the impact case study by Sheppard [8], the research impact included encouraging scientists to look beyond the current aesthetic of digital images and helping patients to communicate their experiences individually and collectively through artist exhibitions. Obviously, citation measures in these cases would have been unable to capture such qualitative effects.

On a macro-level of research impact, three factors will determine the overall impact of research: the geographical and disciplinary extent to which research output has been diffused, the adoption rate, and the societal benefits. The units and implications of this statement can be more formally expressed with the equation[1]:

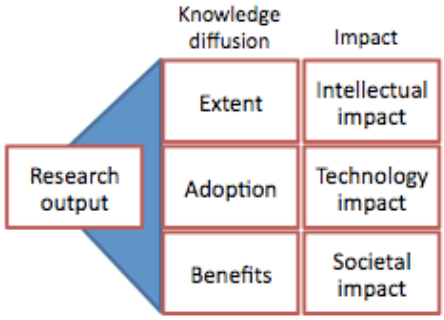


Figure 2. The research impact framework

$$I = E \times A \times B \qquad \text{[Equation 1]}$$

where the overall impact I is defined as the product of the extent (E) of knowledge diffusion, in which citation data can be utilized to analyze the rate and scope of knowledge diffusion, the rate of adoption (A) as represented by the proportion of intellectual property that has been licensed or purchased among all produced, and the benefits (B) to society in both quantitative and qualitative terms.

Figure 2 visualizes Equation 1 as a framework for evaluating the impact of research, which was created with information science in mind in particular. Through knowledge diffusion, various types of research output make impact at intellectual, technological, and societal levels. At each stage of diffusion, the kinds of impact that research output has are assumed to be different: the geographical and disciplinary extent of knowledge diffusion produces mostly intellectual impact, the adoption of inventions or innovations makes impact mainly on technology and implementation, and benefits to communities and society would impact the society as a whole in much broader and deeper ways. While these assumptions are yet to be proved, differentiating between the kinds of impact would help us clarify what measures are appropriate for assessing which types of impact and thus address the long-time concerns on the validity and reliability of impact measures.

2.1 Extent

So far citation counts have been the most widely used measure for impact at either individual or collective levels for papers, journals, or institutions (Given its wide use in science research performance and impact assessment, it exemplifies an excellent case in intellectual impact, to say the least). For reasons mentioned in the introduction section, citation measures are not without problems: they may be inaccurate, misleading for interpretation; non-experts may use them inappropriately in evaluating research quality and

[1] The equation idea comes from Parker, I.M. et al. (1999). Impact: toward a framework for understanding the ecological effects of invaders. *Biological Invasions*, 1: 3-19. While the variables are different in this equation from the one in Parker et al.'s article, the way the authors describe the impact of biological invaders on ecological systems and economy helped formulate the equation in my paper.

performance [10]. While citation data has its advantages in research evaluation, its role lies mainly in measuring the extent of knowledge diffusion only in formal scholarly communication, rather than measuring the whole spectrum of research impact as many bibliometric studies have taken for granted. It translates the quantity of citations into one type of impact—the intellectual impact—and such impact is only in proxy. As for the qualitative aspects—whether the research had an impact on others in terms of methods, hypotheses, theories, or experiments, or whether it served a supportive role or otherwise for the citing work, the assessment must go beyond citation data and quantitative methods to obtain a holistic picture of research impact.

2.2 Adoption

Adoption is a term used in the diffusion of innovations theory. Diffusion is defined as "the process by which an innovation is communicated through certain channels over time among the members of a social system" [7, p.5]. While diffusion research centers on how (diffusion process) and why (adopter studies) innovations are diffused, impact assessment is targeted in the outcomes of such processes and their impact. Some of the outcomes from adoption are countable, e.g., number of license agreements signed, number of patents applied or approved, or academic-industrial research partnerships established. Others, however, are not so direct and easily countable. The impact of adoption is not always immediately obvious. A good example would be the involvement of science librarians in managing datasets for scientists. While science metadata and datasets management result from practice-based research, the adoption of science metadata standards and best practices in supporting eScience is a typical stage of diffusion. The impact of this adoption may not be immediately clear in the short term nor easily measurable within the information science field, but anecdotal evidence has shown that such an adoption is changing the curriculum structure and guiding principles for educational programs in some institutions. A systematic assessment of such an impact would require data collection from outside of research field itself.

Here again disciplinary differences affect the kinds of impact resulted from adoption. For example, the impact of an adoption of biological materials may bring benefits in clinical trials that show immediately in patient's condition change, while that of metadata best practice guidelines may not be so obvious nor direct because the change in metadata quality and search performance cannot be known until metadata quality is inspected and search results are analyzed against the queries.

2.3 Benefits

If E and A are relatively straightforward to quantify, benefits B is not. This aspect of research impact has always been a challenge. Macro-level measures such as proportion of Gross Domestic Product (GDP) increase due to Research & Development (R&D) spending signals the contribution of R&D activities to a country's economic growth, but how can we quantify the benefits that resulted from information science research at both macro- and meso-levels? In other words, how does information science research contribute to the improvement of productivity and quality of life?

Measuring benefits of information science research needs to consider two important factors: scope and type. The scope factor includes societal, organizational, and individual levels. A research output may benefit individuals in one way but this benefit may be translated into other types for an organization or community. A good example would be the development of institutional repositories (IRs). While theoretical and application research on this subject help build effective information systems in the form of IRs, the benefits for researchers in the institutional community may be measured by the time saved in managing their research artifacts (including data among other research products) from avoiding technical hassles. Although the data of benefits may be difficult to collect, it is not impossible to capture and the chain benefit change at individual, institutional, and societal levels would make such data collection worthwhile. Research on the B segment in Equation 1 would be the most challenging yet perhaps most revealing way to assess the impact of information science research.

3. How Is Impact Measured Currently?

A large number of measures have been used to assess research impact. Bollen et al. [3] analyzed 39 existing and proposed measures of scholarly impact and concluded that no single indicator can be used alone to measure scholarly impact. Table 1 summarizes the measures listed in Bollen et al. from a methodological perspective.

Citation measures have functions of ranking, measuring citedness and relatedness of formal publications, and supplying data for mathematical models of research achievements. The popular use of citation data did not make it more effective or accurate in evaluating impact. Bollen et al.'s research reveals that usage measures show a greater reliability than citation measures in general and "Usage Closeness centrality is positioned closest to all other measures" [3]. One point worth making here is that, although Bollen et al. call all the measures in Table 1 as "scientific impact measures," they are all citation-based measures and have the inherent limitations for getting the whole picture of impact.

Table 1. Existing measures for scholarly impact (compiled according to Bollen et al., [3])

Function of measures	Type	
	Citation	Usage
Ranking	Scimago Journal Rank, PageRank, Y-factor	PageRank,
Citedness	Cites per doc, Journal Impact Factor, Scimago Total Cites, Journal Cite Probability	Journal Use Probability, Usage Impact Factor
Relation	Closeness centrality, out-degree centrality, degree centrality, in-degree centrality, betweenness centrality	Closeness centrality, degree centrality, in-degree centrality, betweenness centrality, out-degree centrality
Index	Immediacy index, H-index, citation half-life	

4. Conclusion

Needless to say, assessing the impact of research requires more than citation data. Although bibliometric studies have developed numerous measures for research impact, or "scientific impact," they are all citation-based measures and thus unavoidably limited by the inherent problems due to these constraints. This paper analyzed the misconception about what research impact is and what exactly citation data measures or cannot measure for impact. The framework of impact assessment in this paper, though still in its preliminary form, offers a more holistic view on the composite of research impact.

The three components of research impact, Extent of (intellectual) impact, Adoption of technology and practices, and Benefits to society, emphasize the diffusion of knowledge processes and outcomes. How to operationalize the measures in the impact model will need further research.

While current assessment of research impact relies largely on empirical data and measures, there is a lack of theory and formal (rational) models for evaluating the impact for the information science discipline. The framework offered in this paper is an attempt to fill this blank. The next step will be to further refine the detail of the framework and gather more empirical evidence to rationalize it into a model and theory.

5. Works Cited

[1] Bordons M, Fernandez MT, Gomez I. 2002. Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country. *Scientometrics* , 195-206.

[2] Becker Medical Library. 2009. Becker Medical Library Model for Assessment of Research Impact. http://becker.wustl.edu/impact/assessment/model.html

[3] Bollen J, Van de Sompel H, Hagberg A, Chute R, 2009. A Principal Component Analysis of 39 Scientific Impact Measures. PLoS ONE 4(6): e6022. doi:10.1371/journal.pone.0006022

[4] Duryea M, Hochman M, Parfitt A. 2007. Measuring the impact of research. *Research Global*, February: 8-9, 27. http://www.atn.edu.au/docs/Research%20Global%20-%20Measuring%20the%20impact%20of%20research.pdf.

[5] Gagolewski M, Grezegorzewski P. 2008. A geometric approach to the construction of scientific impact indices. . *Scientometrics* , DOI: 10.1007/s11192-008-2253-y.

[6] Oswald AJ. 2009. A suggested method for the measurement of world-leading research (illustrated with data on economics). *Scientometrics* , DOI 10.1007/s11192-009-0087-x.

[7] Rogers EM. 2003. *Diffusion of Innovations.* New York: Free Press. 5th ed.

[8] Sheppard D. 2007. Social impact of artist exhibitions: Two case studies. The Arts and humanities Research Council (UK). http://www.ahrc.ac.uk/About/Publications/Documents/Social%20Impact%20Exhibitions%20Web.pdf.

[9] van Leeuwen TN, Moed,HF, Tijssen RJ, Visser MS, van Raan, AF. 2001. Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics* , 51: 335-346.

[10] van Raan AF. 2005. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. . *Scientometrics* , 62(1), 133-143.

[11] Weingart P. 2005. Impact of bibliometrics upon the science system: Inadvertent consequences? . *Scientometrics* , 62(1), 117-131.

(Measuring Research Impact) "I Stay Away from the Unknown, I Guess," Measuring Impact and Understanding Critical Factors for Millennial Generation and Adult Non-Users of Virtual Reference Services

Marie L. Radford
Rutgers University
School of Communication and Information
732-932-7500 x8233
mradford@rutgers.edu

Lynn Silipigni Connaway
OCLC
OCLC Research
6565 Kilgour Place
Dublin, OH 43017
(303) 246-3623
connawal@oclc.org

1. INTRODUCTION

Although research on Virtual Reference Service (VRS) users has proliferated since its beginnings in 1999, a negligible amount is known about non-users and the reasons why they do not select VRS for their information needs. The international study "Seeking Synchronicity: Evaluating Virtual Reference Services from User, Non-User, & Librarian Perspectives"[i] investigated critical factors in selection, use, and satisfaction of synchronous, live chat services. The project involved several data collection techniques (transcript analysis, focus group interviews, online surveys, individual interviews) using quantitative and qualitative methodologies. The project's four phases involved: focus group interviews; online surveys; telephone interviews with VRS users, non-users, and librarians; and analysis of 850 QuestionPoint[ii] live chat transcripts. This paper reports results from online surveys and telephone interviews for non-users.

Theoretical frameworks from [1] and [2] as applied to face-to-face (FtF) [3,4] and chat [5,6] reference encounters were used to develop research questions and to guide survey instrument development and data analysis. These research questions also developed from the project's focus group and transcript analysis results and from the literature review:

- What are VRS non-users' communication and information-seeking preferences?
- What factors would influence non-users decisions to use VRS?
- What are critical factors in successful reference encounters?
- What is the relative importance of getting an information/answer vs. how one is treated in determining success?

2. DATA COLLECTION AND ANALYSIS

Online survey and telephone interview questions emerged from the analysis of the focus group interviews and the chat transcripts. Non-users (those who had never used VRS, but may be using Instant Messaging (IM) or chat for social or business purposes and may also be users of physical or digital libraries), were recruited for both the online survey and telephone interviews through a variety of methods including university email listservs and posting of flyers.

VRS non-users completed 184 online surveys and 107 telephone interviews featuring quantitative and qualitative questions. Data was collected from 6/2007 to 3/2008. The team used descriptive statistics for quantitative data and grounded theme analyses [7] and the Critical Incident Technique (CIT) [8] for qualitative data.

3. DEMOGRAPHICS FOR ONLINE SURVEYS AND TELEPHONE INTERVIEWS

The majority of online survey and phone interview participants were Caucasian, female, used public libraries, and suburban libraries, but had not tried live chat VRS. Focus group and transcript analysis revealed generational differences, so data for Millennial generation[iii] (12 - 28 years old) respondents (aka Generation X, Net Gen) was compared to older adults (29+)[iv]. (See Tables 1 and 2).

Table 1: Millennial Demographics Online Surveys & Telephone Interviews (N=195)

	Total	%
Gender		
Female	124	64 %
Male	71	36 %
Age		
12-14	23	12 %
15-18	59	30 %
19-28	113	58 %
Ethnicity		
African American	16	8%
Asian/Pacific Islander	37	19 %
Caucasian	127	65 %
Hispanic/Latino	8	4%
Native American	2	1%
N/A	1	1%
Other	4	2%
Types of Library		
Academic	35	18 %
Public	90	46 %
School	70	36 %
Location		
Urban	73	37 %
Rural	12	6%
Suburban	110	56 %

Table 2: Older Adult Demographics Online Surveys & Telephone Interviews (N=95)

	Total	%
Gender		
Female	72	76 %
Male	23	24 %
Age		
29-35	20	21 %
36-45	26	27 %
46-55	27	28 %
56-65	15	16 %
65+	7	7%
Ethnicity		
African American	6	6%
Asian/Pacific Islander	3	3%
Caucasian	79	83 %
Native American	1	1%
Other	4	4%
N/A	2	2%
Type of Library		
Academic	17	18 %
Public	76	80 %
School	1	1%
Special	1	1%
Location		
Rural	6	6%
Suburban	56	59 %
Urban	33	35 %

4. QUANTITATIVE RESULTS ONLINE SURVEY

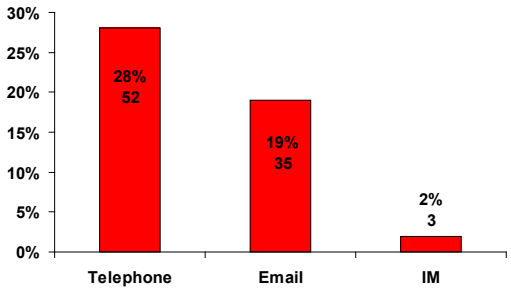
4.1 Online Surveys Demographics

Of 184 online survey respondents, 66% (122) were Millennials and 34% (62) were older adults. As noted above, the majority for both groups were female, Caucasian, and used public libraries and suburban libraries.

4.2 Reference Experience

In addition to FtF interaction, participants reported use of other modes with 28% (52) having used the telephone, 19% (35) email, and 2% (3) IM reference. Phone reference was used by 78% (95) of Millennials versus 60% (27) of adults. (See Figure 1.)

Figure 1: Experience with Reference Modes: Non-User Online Survey (N=184)



When asked about their preferences 81% (N=50) of adults and 71% (N=87) of Millennials were partial to FtF, as illustrated below:

> "Most recently I wanted to read about an actor that I really enjoy. I got into a conversation with the librarian about him and she was able to locate a number of books on him, including his memoirs…she suggested that I check with videos to see what might be available and again she assisted in finding at least half dozen that I was able to borrow…this experience gave me a great weekend in addition to some very enjoyable reading material … In this particular circumstance having a face-to-face enabled us to share a more personable and friendly exchange of information."[v] (Adult)

> "I used face to face format because I think it is more direct and you are more likely to get an answer quicker, plus you are right there so you can learn things like about different reference websites. It did help by experience to be successful I feel that if you talk to someone face to face it is more personal and more helpful." (Millennial)

As Figure 2 shows, 49% (60) of Millennials enjoyed FtF above email (27%, 33), phone (12%, 14), or texting (12%,15) for reference as exemplified below:

> "I believe the face-to-face format helped my experience to be successful. This is because the interaction was far more personal, I was able to clearly state my question and get immediate feedback or answers. She was able to clarify what it was that I was looking for and was there waiting for me to come back if I had any trouble finding what I needed once she had given me the locations of what I was looking for." (Millennial)

> "I have nothing truly against chat reference services, so I may use it in the future, but I will probably always rely on the face-to-face services as my main form of information seeking." (Millennial)

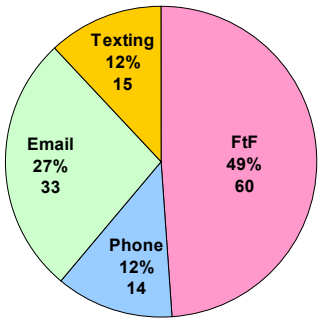


Figure 2: FtF Preferred by Millennials: Non-user Online Survey (N=122)

"I most enjoy using"

4.3 Email Reference Least Intimidating

Millennials most enjoyed FtF reference, but 51% (62) reported being less intimidated by email, followed by FtF (20%, 24), texting (17%, 21), and telephone (12%, 15) (see Figure 3).

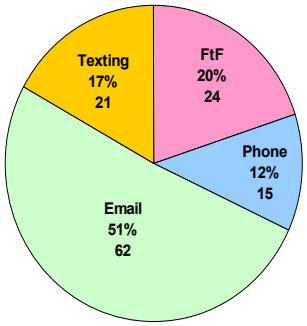


Figure 3: Email Least Intimidating to Millennials: Online Survey (N=122)

"I am least intimidated by"

4.4. Convenience

Responding to the statement: "The library is convenient," 76 (87%) of the 87 Millennials and 40 (78%) of the 51 adults answered positively. However, some participants commented that online sources are more convenient:

> "Going to an actual library would be my last resort. I can get everything that I can get at a library and more online, and I don't have to go anywhere. I guess that's what influences it." (Millennial)

> "If I needed to find out anything I would usually go first online. That's really my main source of information, because it's really convenient. The fact that it is really convenient: I don't have to go out of my room to find the source, just go online and try and can just hit enter, and it's really convenient." (Millennial)

4.5 Remote Access

Reacting to the statement, "Remote access is important," 95% (39) of the Millennials and 85% (11) of the adults answered positively. This comment is typical:

> "Because I am on the computer a lot anyway and I wouldn't have to leave my house to physically go to the library to get some answers. It is just as easy for me to formulate my questions online and it would save me time." (Millennial)

4.6 Personal Relationships

Personal relationships with librarians were considered to be more important for adults 43% (22) (N=51) than Millennials 24% (21) (N=86). Adults 51% (26) were more likely to value interactions with specific librarians than Millennials 24% (24). Illustrative comments follow:

> "I never want a computer interface to replace face to face contact with a person. In this day in age, it might be more convenient to jump on the web to get the information you need, but I think you potientially missing connections a library creates. In my business experience, email can only take you so far. Conference calls and face to face meeting provide the connections that emails can often destroy. Service should never be an either/or situation. Personal contact and computer interface connections should exist together." (Adult)

> "I was at my town's public library, and there is this one lady who works in the Reference department…One time, I needed some books and she looked it up for on the computer and said that this library didn't have the book, but another library did. She wrote down the information for me…This lady is very helpful, she makes you feel like she actually wants to help you and goes onto the computer, looks up the subject you're working on, and finds the books for you, and if the book(s) isn't there, she'll help you request it. Other librarians don't offer this same help." (Millennial)

4.7 Friendliness/Politeness

69% (28) of the Millennials (N=41), and 29% (4) of older adults (N=14), valued the FtF librarians' friendliness and politeness. A sample statement from a Millennial:

> "I liked the one-on-one interaction, which enabled me to have my specific questions answered on the spot. The librarian was able to address my specific needs with practical, useful information. She was friendly and appeared genuinely glad to be helping me. I think the face-to-face format did help, since it was a relaxed meeting. I was comfortable with the librarian, so I was comfortable asking questions. The in-person meeting was necessary to help me learn how to locate articles on microfiche and how to use the equipment." (Millennial)

One Millennial shared an experience with an unfriendly librarian, contrasting her behavior with that of another librarian.

> "It was awhile ago, but I asked the reference librarian where to find books relating to a certain topic I was studying in school at the time and she just kind of said "over there" and pointed…She did not seem engaged or interested in truly helping me find the books and didn't really care that I never found them, I was wondering all over looking and she just sat there…I doubt she would have been any more helpful in another format and I don't care if she would have been great at texting etc. because at the time, I was there looking for books and just wanted to know specifically where they were located. She did not seem interested in helping me, let alone exhaust all of her means for doing so.

> Ever since then, I usually avoid that person and go to the one who has helped me successfully." (Millennial)

4.8 Reasons for Not Choosing VRS

Reflecting the Millennial's high level of comfort in the IM environment, 35% (N=43) of 122 Millennials and 53% (N=33) of 62 older adults agreed with the statement, "Chat reference might be too complicated." Adults commented on why they might/might not try VRS:

> "I most likely will not use this service. Computers were not taught in High School when I graduated in 1972, I have only had a computer and used email since 2005, I have never used a chat room or service." (Adult)

> "If they had classes at the Library and showed me how to do it I might try it. My daughter usually shows me everything I need. But I really like going to the Library and talking with someone in person, so I probably wouldn't use the service even if I knew how to use it." (Adult)

More adults (35% 22) were concerned that their typing was not adequate for VR compared to 16% (19) of Millennials. However, the younger cohort (29%, 35) was more concerned that their questions might annoy the librarian and some worried about "bothering" the librarians. One explained why:

> "The librarian I asked seemed too occupied with other matters to pay any attention to my question, and she made me feel stupid and intrusive for even asking her such a thing." (Millennial)

Some did not believe a librarian could help, others did not know VRS existed:

> "I do not see myself using chat reference services because in the absence of having a reference librarian help me locate an appropriate or required source, I have friends in the LIS discipline with exemplary reference/research skills who could help me. Additionally, because I am in research, I have cultivated my own knowledge base of where/how to track down information. The only time I could ever imagine using chat reference is if I were incapacitated or unable to physically be in a library or if I were unable to reach one of my LIS colleagues. Otherwise, I see myself as a self-sufficient researcher who relies on her own social network and knowledge to locate reference material." (Adult)

> "I've never used this type of service and never knew it was available – that's probably why I never tried it. Also, in my everyday life I don't run across the need to research something in depth (the internet usually has enough information) so I don't really have a need to chat with someone for reference help." (Millennial)

Both cohorts did not use VRS because of satisfaction with other information sources (e.g., family, friends, colleagues, teachers, the Web):

> "I choose to go FTF because I'm not lazy, and I can get a more accurate answer FTF not on chat reference. And I can be there to get the books I want, and not waste a librarian's time on the computer." (Millennial)

> "Why use VRS when phone, face-to-face, or even e-mail could be more convenient?" (Millennial)

> "I don't know how to access computer library service. When I need to look something up I use Google." (Adult)

4.9 Reasons to Use VRS

Non-users thought they would try VRS if they could receive information quickly and around the clock, 24/7/365:

> "If it is available 24/7 I'll try it." (Millennial)

5. QUANTITATIVE RESULTS TELEPHONE INTERVIEWS

5.1 Demographics

Millennials gave 69% (73) of 107 telephone interviews, 31% (33) were older adults, and one gave no age. As above, the majority for both groups was female, Caucasian, used public libraries, and libraries in suburban areas. Most Millennials were aged 19 to 28 and adults were 29 to 35.

5.2 FtF Preferred

The telephone had never been used for reference by 76% (81) of participants, 74% (79) had not used email, and 94% (101) had not used IM reference. Twenty-four percent (26) preferred FtF reference and complimented librarians:

> "[T]hey kind of know, like, almost 'every single book' in the library, or at least they know how to use the computer to find the book itself. And if I need help on some kind of information, they know stuff about what kind of books I need to use." (Millennials)

> "I think reference librarians are up there with fire fighters as my heroes." (Adult)

5.3 Reasons for Not Choosing VRS

When asked why they did not use VRS, 82% (24) of 34 interviewees responded that they were unaware that it existed.

> "I think I am unfamiliar with it. I don't have info or familiarity with it. I stray away from the unknown I guess." (Millennial)

When asked what alternatives to the library they chose, 43% (45) of participants said they used the Web. Specifically, 45% (33) of Millennials and 28% (20) of adults used the Web for "personal convenience." Google was mentioned by 15% (11) of Millennials and 3% (2) of adults:

> "Say I have physics homework, I wouldn't use the Internet to find information. I would find a person to help me understand it. But if I have a topic to research, I probably wouldn't go to a person, I would go straight to Google or Yahoo and research away." (Millennial)

5.4 Reasons to Use VRS

When asked what might convince them to ask for help from VRS, convenience again was mentioned, including 24/7 access to librarians. Thirty percent (32) of respondents cited immediate answers and 17% (18) appreciated home access.

> "It would be convenient, because if I was sitting at a computer and I could ask a question and they would answer immediately… that would be good.. Convenience is why I do something as opposed to something else." (Millennial)

6. QUALITATIVE DATA ANALYSIS – CRITICAL INCIDENTS (CIs)

Qualitative data were analyzed with Flanagan's CIT [8], a qualitative method which is used to evaluate programs or services; including reference encounters (name removed). CIT is "often used to study effective and ineffective human behavior" [9] as it allows critical factors to emerge rather than be imposed.

Online survey and telephone interview instruments included two CI questions, which asked participants to "think about one experience" in which he/she felt a positive result and in which a negative result was achieved after seeking reference services. Then they were asked to describe each interaction and to identify factors that made it successful/unsuccessful.

The CIs were sorted into content themes following the constant comparative method [6,10]. Relational theory provides the analytical framework. It posits that every message has dual dimensions – both content (information) and relational (interpersonal) [1]. Emerging themes were expanded and merged into the Critical Incident Coding Scheme developed by [9, 11], for FtF reference encounters and VR encounters [5]. Content themes relate to interactions that focus on the information sought, the degree to which participants perceived that librarians interacted knowledgeably. The relational themes outline personal aspects, including attitude, relationship quality, and approachability.

7. QUALITATIVE RESULTS ONLINE SURVEYS

7.1 Positive CIs

A total of 154 (84%) positive CIs were collected from 184 online surveys. 108 (70%) of these were described by Millennials. Participants attributed success to information delivery/retrieval (50%, 54), the librarians' positive attitude towards individual and task (36%, 39), location of specific resources (20%, 22), and answering questions (14%, 15). A number of themes are revealed in positive CIs, as shown in below examples:

> "I asked the librarian where the murder mystery books were located, she was kind of busy checking in books, but still took the time to answer my question. She put down what she was doing and she walked me to the correct section, instead of just pointing me that way." (Millennial)

> "I was looking for books on theoretical physics. My question was, 'What would be the latest and most comprehensive book on quantum electrodynamics?' I felt the encounter was successful because the librarian apparently had a background in physics. He was up to date in his physics knowledge and was aware of the latest books." (Millennial)

7.2 Negative CIs

Negative CIs were provided by 99 (54%) respondents, of which 75% (74) were Millennials. Unsuccessful experiences were attributed to librarians who impeded information delivery/retrieval (64%, 47), had a negative attitude towards task (47%, 35), or were slow in providing answers (11%, 8). Missing resources (12%, 9) were also reported as negative. Participants described numerous barriers in their negative CIs.

> "At one point when I had asked for service from a librarian, it seems like her attitude downplayed my intelligence and because she was older and wiser my question wasn't of importance. The attitude of the librarian was not friendly and welcoming and I sort of

felt embarrassed after asking for assistance." (Millennial)

"I tried to explain that I wasn't interested in doing a general search on my topic, but that instead I needed this specific article, but she never really listened, and instead I ended up wasting a significant amount of time. The librarian was so overzealous with helping me that she lost sight of what I actually needed, which in this case was quite limited in scope, a specific item." (Adult)

8. QUALITATIVE RESULTS TELEPHONE INTERVIEWS

8.1 Positive CIs

Telephone interviews yielded 122 positive CIs. Seventy-nine percent (84) praised librarian's personal characteristics with 50% (61) of these crediting the librarians' knowledge/skills. Information aspects were primary in 49% (52) of CIs while 46% (49) simply found the librarians helpful. Some CIs related directly to FtF communication that included nonverbal communication cues. These comments exemplify positive CIs:

"Well, at my library they are all very approachable: they are just sitting at a desk waiting to help, kind of not judging i guess." (Millennial)

"I'm a elementary teacher, and my most recent experiences have been with children's librarians so I think it's mostly their knowledge, not just of the library catalog and ways of manipulating the catalog but what material is out there … it's very beneficial." (Adult)

8.2 Negative CIs

Telephone interviews yielded 112 negative CIs. The greatest proportion (23, 21%) centered on librarian's characteristics, including being unapproachable or lacking knowledge. Others (13, 12%) thought librarians' were not helpful, did not answer the question, responded slowly, or gave too simple a response. Examples of Negative CIs include:

"I felt the FtF helped to make it successful. I was in front of her and the information was straightforward and she looked me face to face. In an email she would not be in front of me and I would not know if she was being truthful." (Millennial)

9. DISCUSSION AND IMPLICATIONS

Although accuracy and correct answers and the delivery of specific content were reported as the most important factors of successful reference interactions, non-users of VRS also value librarians who are knowledgeable about information sources and systems, display a positive attitude, and demonstrate good communication skills.

Non-users of VRS are not aware that the service is available. Results from both interviews and surveys reveal that they consider convenience to be a major factor when choosing how to get their information. Respondents prefer to interface with friendly librarians and to develop ongoing relationships with them. The majority used FtF as well as telephone and email reference services. They found email reference to be the least intimidating mode of communicating with a librarian for a reference query, but most preferred FtF because they felt the interaction was more personal, more efficient, and enabled them to better communicate with the librarian. Although most preferred FtF reference services and believed the library is convenient, some said online sources are more convenient than physical library materials because of remote access.

Many of the non-users did not believe a librarian could help them or know that VRS was available. Both Millennials and adults were satisfied with other information sources; therefore, did not need to use VRS. Human resources, such as family, friends, teachers, and colleagues were identified as prime information sources. The Web was identified as an alternative for the library and was used for "personal convenience." The non-users might use VRS if it were available 24/7 and if they could receive information quickly.

Some differences in communication and information seeking behaviors were found between Millennials and adults. A personal relationship with a librarian was more important to Millennials who also valued the librarians' friendliness and politeness in interpersonal communications more than adults. A greater number of adults than Millennials believed that chat reference would be too complicated; therefore, chose not to use it. The adults were concerned that their typing skills were not adequate to communicate with a librarian via chat. On the other hand, Millennials were more concerned than the adults that their questions would annoy or bother the librarians.

10. CONCLUSION

In these tight budget times, library service providers must seek to understand non-users of the library to better meet their particular needs and preferences. The above findings have numerous implications for librarians who are involved in making decisions that will have a positive effect on sustainable VRS in the future. Results can be used for system development, improving VR practice, and for theory development. The voices of the little-studied non-user population provide powerful evidence that libraries need to step up marketing of these services. Once these

potential users are aware that the services exist, that virtual librarians are accurate as well as friendly, and are knowledgeable and technically competent, these non-users can be enticed to view virtual services as a viable and attractive alternative to FtF, telephone, or email reference.

11. ACKNOWLEDGEMENTS

The authors would like to acknowledge the help of Erin Hood and Timothy Dickey of OCLC and Jocelyn DeAngelis Williams of Rutgers University.

12. REFERENCES

[1] Watzlawick, P., J. Beavin, and D. D. Jackson. 1967. Pragmatics of Human Communication. Norton.

[2] Goffman, E. 1967. Interaction Ritual, Essays on Face-to-Face Behavior. Doubleday.

[3] Radford, M. L. 1993. Relational aspects of reference interactions: A qualitative investigation of the perceptions of users and librarians in the academic library. PhD diss., Rutgers, The State Univ. of New Jersey.

[4] Radford, M. L. 1999. The Reference encounter: Interpersonal communication in the academic library. ACRL, A Division of the American Library Association

[5] Radford, M. L. 2006. Encountering Virtual Users: A Qualitative Investigation of Interpersonal Communication in Chat Reference. Journal of the American Society for Information Science and Technology. 57, 1046-59.

[6] Radford, M. L. 2006. Interpersonal Communication in Chat Reference: Encounters with Rude and Impatient Users. In: The Virtual Reference Desk: Creating a Reference Future, ed. R. D. Lankes, E. Abels, M. White, and S. N. Haque, 41-73. Neal-Schuman.

[7] Charmaz, K. 2006. Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. Sage.

[8] Flanagan, J. C. 1954. The Critical Incident Technique. Psychological Bulletin. 51, 327-58.

[9] Ozkaramanli, E. 2005. Librarians' Perceptions of Quality Digital Reference Services By Means of Critical Incidents. Doctoral Thesis. University of Pittsburgh.

[10] Strauss, A., and J. Corbin. 1998. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2nd ed. Sage.

[11] Radford, M. L. 2006. The Critical Incident Technique and the Qualitative Evaluation of the Connecting Libraries and Schools Project. Library Trends 54(1), 46-64.

[12] Radford, M. L., and L. S. Connaway. 2005-2008. Seeking Synchronicity: Evaluating Virtual Reference Services from User, Non-user, and Librarian Perspectives. Funded by the Institute of Museum and Library Services (IMLS). http://www.oclc.org/research/projects/synchronicity.

[13] Radford, M. L., and L. S. Connaway. 2007. "Screenagers" and Live Chat Reference: Living Up to the Promise. Scan. 26(1), 31-9.

[14] Connaway, L. S., M. L. Radford, T. J. Dickey, J. D. Williams, P. C. Confer. 2008. Sense-making and Synchronicity: Information-Seeking Behaviors of Millennials and Baby Boomers. Libri. 58, 123-35.

[15] Connaway, L. S. 2008. Make Room for the Millennials. NextSpace. 10, 18-9. http://www.oclc.org/nextspace/010/research.htm

NOTES

[i] This project was funded by a grant from the Institute of Museum and Library Services (IMLS) and in-kind contributions from Rutgers, The State University of New Jersey, and OCLC Online Computer Library Center, Inc., 2005. Grant website is http://www.oclc.org/research/activities/synchronicity/default.htm [12]

[ii] QuestionPoint "provides libraries with tools to interact with users in multiple ways, using both chat and email." OCLC Web Site http://www.oclc.org/us/en/questionpoint/default.htm

[iii] The authors have provided in-depth discussions of the characteristics and behaviors of the Millennial Generation [13-15].

[iv] One respondent did not reveal their age and is not included in Millennial/Adult counts so N=290, otherwise respondents total N=291.

[v] Grammatical errors have not been corrected, although minor spelling errors have been corrected in quotations from participants..

The Ontology of Tags

David J. Saab

College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802

dsaab@ist.psu.edu

ABSTRACT

Social bookmarking sites such as Flickr, del.icio.us, and CiteULike have adopted folksonomic systems where users tag entities with keywords. These tagging systems replace traditional taxonomic systems that employ hierarchical categorization schemes. While there are some differences in how these tagging systems are constructed, e.g., as broad or narrow folksonomies, there has been confusion as to whether tagging constitutes a collaborative activity or a collective one. The distinction between collaborative and collective influences the theoretical assumptions upon which research is conducted. Researchers have adopted a semiotic theoretical perspective as an avenue for discerning emergent semantics of folksonomies. If tagging systems are to be useful to social media or semantic technologies, if we are to indeed discern the semantics emergent from folksonomies, then we need to understand the ontology of tags. This paper examines some of the fundamental ontological assumptions regarding tagging and folksonomies.

Categories and Subject Descriptors

H.1.1 [**Information Systems**]: Models and Principles—Systems and Information Theory (Value of information); H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing Methods; H.3.5 [**Online Information Services**]: Data sharing

General Terms

Theory, Design

Keywords

Folksonomies, tags, tagging, cultural, schemas, ontology, semantics, Heidegger

1 TAGS IN FOLKSONOMIES

Folksonomies is a term coined by Vander Wal [39] to refer to the "result of personal free tagging of information and objects for one's own retrieval." Tagging happens in a social environment and is done by individuals consuming information. Folksonomies are similar to taxonomies in that both use keywords to describe information or objects within a domain. The term, folksonomies, is a combination of folk and taxonomy, which is a bit of a misnomer since folksonomies lack the one critical characteristic of all taxonomies—hierarchy. Vander Wal considers folksonomies to be complements to taxonomies rather than replacements for them. Shirky [30] makes the case for the use of folksonomies rather than rigidly structured categorization schemes. He raises the issue of the "information explosion" as a primary force in the shift from standard classification schemes, such as librarians use, to tagging and folksonomies that are non-hierarchical user-developed classification systems.

Tags are generated by individuals for their personal use, to be able to retrieve information and/or objects quickly and in a way that conforms to their understanding of the entity. Social bookmarking sites as Flickr, del.icio.us, and CiteULike have incorporated the use of tags as way for users to retrieve photos, URLs, and citations in a way that is personally meaningful and which doesn't require learning taxonomies constructed by professionals. Users employ their own vocabulary, which has meaning specific to them. It is these meaningful associations expressed as tags that enable faster and more direct recall of the object because they act as representations for the way we think [14].

However, when researchers study the folksonomies of del.icio.us, for example, they group together all of the tags created by all users for a particular resource as if it was representative of a single perspective. They do not attempt to make any distinctions between users, often because they have no identifiable or discrete information about them. This approach is problematic insofar as a single individual can effortlessly switch their perspectives based on their identity and create tags for the same phenomenon based in different, sometimes conflicting, identities. For example, we can imagine that a person who is a hunter might tag a geographic area within a GIS as “exciting” or a web page about weapons as “essential resource.” That same individual using his identity as a father might also tag the same geographic area and web page as “dangerous” and “prohibited,” respectively. The relative simplicity of the tagging concept is transformed into a problem of greater complexity when we begin aggregating tags into tagclouds and broad folksonomies [38] associated with particular perspectives—cultural identities and schemas. Compounding this complexity is the fact that many perspectives exist as part of an individual's cognition, and that the same perspective can be used as an identity for many individuals.

Tags in isolation are not very semantic. A word isolated from the entity it was intended to describe and from the person who created it can mean or refer to many things, and many people may interpret the same tag differently based on their personal histories. In order to make sense of a semantic tag, it is important to understand the perspective from which it is offered. Tags are

ontic signs that serve as indicators to the rich ontological conceptualizations we hold in cognition. Because each individual has a different experiential history, we would expect that their ontological conceptualizations to be unique. Individuals are also members of cultures, and as cultural beings they share many common experiences and articulate them using language. We learn the languages of our parents and communities as children, and share a vocabulary that enables us to express meaning regarding our experiences. Language may be simply words vocalized or written, but intonation, demeanor, time, context, etc. all play into the semantics of the expressed language and facilitate our understanding of others.

The following sections address the issues of shared vocabulary and the semiotic tri-concept relationship between user, tag and entity. The analysis and argument offered towards understanding the issue of emergent semantics of tags will draw upon cultural theory and Heideggerian phenomenology to articulate their ontology.

2 THE CULTURAL NATURE OF TAGS

While folksonomies do not explicitly state the relationships that exist in a conceptualization, the use of tags by users with similar interests tends to converge to a shared vocabulary [1, 7, 19, 25, 40]. Vocabulary convergence is treated as a collaborative activity by researchers, but there is some confusion as to whether sets of tags constitute a *collaborative* activity or a *collective* one [39]. This confusion has implications for how researchers understand folksonomies and their approach to analyzing them. They describe folksonomies as products of collaborative tagging, which is a common characterization in semantic web research [2, 4-6, 19-21, 27, 29]. However, collaborative implies working together towards some goal—that there is active, focused, and agreed upon intent among a group of persons to achieve a specific goal or set of goals [18, 22, 26, 41]. A group agreeing to tag a particular set of resources using an agreed upon vocabulary would be an example of collaborative tagging. Folksonomies are not collaborative in the sense that there are articulated goals towards which the persons creating tags are driving, sans any prior agreement. They are created through a *collective* tagging process, not a collaborative one [39]. Assuming collaboration situates a folksonomy within the confines of a single culture. However, most folksonomies are not so confined. They are open to individuals who have many cultural affiliations and identities, many nationalities and ethnicities, many research domains and spheres of interest [36]. In other words, many cultures, and we can never be certain that the collective set of tags reflect the cultural conceptualization of a particular group.

Culture, as described here, is an emergent phenomenon arising through the interplay of patterns within cognition with patterns extant in the world we inhabit [8, 35]. Schemas, as complex cognitive associations, are *intrapersonal* structures. The objects or events that are manifest outside individual cognition, the entities in the external world, are *extrapersonal* structures. Culture consists of the interplay between the intrapersonal cognitive structures and extrapersonal structures such as systems of signs, infrastructure, environment, social interaction, and so on. The intrapersonal and the extrapersonal are different and distinct, but closely interconnected. They are not isolated from one another, rather separated by a permeable boundary. Culture encompasses both intrapersonal and extrapersonal structures and emerges from the interplay between them. It is through this interplay that we can see that some of the intrapersonal cognitive structures called schemas are shared with others, making them cultural schemas.

The notion of identity and multiplicity of perspectives is important in our understanding how cultural schemas manifest. Individuals can manage multiple identities in the same or multiple contexts. We can shift our perspective effortlessly between national, familial, peer and other identities to make sense of particular phenomena (i.e., frame it in relation to ourselves). The same context, for example, that would be considered "exciting" to "the hunter" might also be "dangerous" to "the parent." Fauconnier and Turner [10] claim that “frames structure our conceptual and social life and, in their most generic and schematic forms, create a basis for grammatical construction.” Words are themselves viewed as constructions, and lexical meaning is an intricate web of connected frames. They also claim that although cognitive framing is reflected and guided by language, it is not inherently linguistic—people manipulate many more frames than for which they have words and constructions. It is the individual's salient, contextualized identity in relation to the phenomena that allows for sense making of the phenomena. When making meaning of a particular phenomenon, individuals will rely upon the cognitive and cultural schemas that are integral parts of their salient, contextualized identities.

The collective nature of folksonomies is indicative of culture only in a very broad sense (e.g., Western culture, English-speaking culture). We should not mistake the tag representation for the underlying ontological conceptualization. A tag is ontic, not ontological, and as such it “*functions both as this definite equipment and as something indicative of the ontological structure of readiness-to-hand, of referential totalities, and of worldhood*” (*Being and Time*, p. 114; H. 83).[1] As an instance of the ontic, it represents an extrapersonal structure. In order for a tag to be considered as part of a cultural phenomenon, it must interact with an intrapersonal schema. Tags will evoke schemas as the individual interacts with them, and it is through this interaction that meaning will emerge. By itself a tag is meaningless and indicative of no particular culture or cultural perspective, per se. When researchers treat tags as if they are ontological, or representative of a single culture’s ontological conceptualization with only minimally recognizable variation, they mistake the collective for the cultural, which is the context where semantics emerge. The assumption is that the conceptualizations brought forth in creating the tag are the same (or only minimally different) for all users who create them.

It is easy to make such an assumption when looking at folksonomies, because they adhere so closely to power law distributions and seem to be remarkably stable. In social bookmarking sites, as entities and phenomena receive more tags, the set of tags as well as the frequency of each tag's use within that set, represents the combined description of that entity by many users [3, 13]. Rather than foster chaotic patterns, the aggregated tags give rise to stable patterns in which the proportions of each tag are nearly fixed. In studying this phenomenon, Golder & Huberman [13] found that after the first

[1] A more thorough explanation of the Heideggerian concepts of *present-at-hand* and *ready-to-hand* is beyond the scope of this brief paper. See (Cerbone 2008) and (Crowell & Malpas 2007) for a more thorough treatment of fundamental concepts in Heidegger’s philosophy.

100 or so bookmarks, each tag's frequency is in nearly fixed proportion of the total frequency of all tags used. They speculate that this stabilization might occur because of imitation and shared knowledge (i.e., a cultural process).

Cultural understanding is expressed through language, and a shared vocabulary is one means by which members of a culture share their understanding of an entity or phenomenon. The shared vocabulary is negotiated over time and evokes shared cultural schemas within an individual's cognition. A shared vocabulary has meaning to the cultural group because the semantics emerge through the evocation of the ontological (i.e., schemas) via the ontic (i.e., tags). The stabilization of tag patterns over time [13] is analogous to the stabilization of cognitive schemas as cultural schemas.

The mere mention of a word is often sufficient to evoke any number of cognitive schemas. As extrapersonal structures, words and language (i.e., tags) serve as social representations that help us identify relationships between images, ideas, objects, and phenomena we encounter in the world [11, 24]. They form the entry points into our complex intrapersonal schemas and rich ontological understanding of experience. What intrapersonal schemas a tag will evoke is dependent upon the cultural context in which it is being experienced [32].

The collective tags of a folksonomy will certainly reflect the dominant cultural schemas of a broad population, but the assumption that collective tags represent a shared conceptualization, interferes with discerning minority cultures, whose schemas may overlap with but are not necessarily entirely consistent with those of the dominant cultural group. In the absence of perspective and cultural identity information about users, folksonomies can be considered as reflections of cultural schemas only for dominant cultural groups and only in the broadest possible sense of "cultural group."

3 THE TRI-CONCEPT RELATIONSHIP

Tagging entities is fundamentally about making sense of that entity. Our experience with those entities allows us to create meaningful conceptual associations for them. Tags may reflect descriptive associations or categorizations of those entities, through which meaning emerges. Semantic web researchers have a strong interest in the semantic dimensions of folksonomies comprised of tags insofar as tags can help structure the information and knowledge available in the vast infosphere of the Web.

Researchers examining the dynamics of tagging systems [6, 12, 13, 20, 31] have settled on a semiotic perspective where tagging is viewed as a tri-concept in which users, resources, and tags are linked. Tags are associated with users who create them and resources to which they refer. A folksonomy is the entirety of a tri-concept tag set—all the tags created by all the users for all resources. Structuring the user-tag-resource relationship as a tri-concept facilitates the analysis of tags with respect to information systems by enabling the application of data mining algorithms to folksonomies [17]. Some researchers focus on identifying the semantic dimensions of folksonomies [1, 4, 5, 34], or understanding their emergent semantics based on this tri-concept model [28] and creating ontologies from folksonomies (or "folksologies" as some researchers put it [23, 33]).

When dealing with the semantic dimensions of tags, issues of polysemy and synonymy reveal themselves [13]. How does one disambiguate among polysemous or synonymous tags? One solution for disambiguating tags is to add a specification to OWL (Web Ontology Language) such that "`<tag> owl: DifferentFrom <tag>`", where the tag is the same lexical unit (e.g., apple) but has different meaning (e.g., fruit vs. computer company) [23]. A complementary suggestion includes the use of "`owl: SameAs`" to merge tags with the same meaning (e.g., *semweb* and *semantic web*). This approach, such as it is, looks promising, but it doesn't easily account for the evolution of the collective lexicon. Also, it would put a burden upon the tagger to specify the "`owl: Relationship`" in a tagging system or it would shift the burden to the ontology revision process, which has its own set of associated problems.

Tags are created at basic, superordinate, and subordinate levels and are related to an individual's interactions with them [37]. There is systematic variation across individuals in what constitutes a basic level, and expertise plays a role in defining the specificity of the level an individual treats as basic:

> *The underlying factor behind this variation may be that basic levels vary in specificity to the degree that such specificity makes a difference in the lives of the individual…. Like variations in expertise, variations in other social or cultural categories likely yield variations in basic levels.[13]*

Tags do need to include the perspective of the tagger in order for semantics to emerge, but recreating the category problem by specifying `DifferentFrom` and `SameAs` relationships only addresses the ontic side of the equation. In order to address the ontological, our understanding of the user as part of the tri-concept relationship must not neglect his cultural perspectives and identities when trying to discern the semantics of particular tag sets. We must consider meaning-making, which is a cultural activity, as a multifaceted process, where semantics emerge through a process of interaction, construction and communication [34]. Interaction involves tasks and activities that generate the need for new meanings based on our being-in-the-world. Construction involves the imposition of "new categories" that are not so-called natural categories in the Aristotelian sense but rather, categories that are based on features that guide retrieval. Communication is negotiated through an alignment of "external tokens" (ontic tags) associated with categories (ontological conceptualizations). There are no "pregiven conventions" or constraints to the communication of categories. "Communication is crucial, because it is the motor for testing the concepts' adequacy and for pushing the development of new concepts when there are misunderstandings of task failures" [34].

Interpretation results from the mutual adjustment of the explicit and implicit content of an utterance. An exhaustive, one-to-one mapping between concepts and words is quite implausible. An interpretation that does not match exactly the intent is not a failure of communication, rather "an illusion of the code theory that communication aims at the duplication of meanings" [32]. Communication succeeds despite semantic discrepancies because the words used in a given situation points the hearer in the direction intended by the speaker. It does not matter whether or not a word linguistically encodes a full-fledged concept, and, if so, whether it encodes the same concept for both speaker and

hearer. Words are used as pointers to contextually intended senses; utterances are merely pieces of evidence of the speaker's intention. We need to know who the speaker is, their identity, in order to interpret the perspective from which the tag originates. The fact that the interpretation of tags is not exact reflects the real-world experience of communication and knowledge sharing and the need for an interactive, hermeneutic discourse to achieve understanding.

Meaning making is a hermeneutic process. If research on the emergent semantics of folksonomies is to be successful, it must incorporate the hermeneutic process of meaning making as part of the tri-concept relationship. The hermeneutic process with respect to the creation and analysis of tags is a process of understanding whereby tags are generated as ontic signs that point to ontological conceptualizations.

This ontic-ontological distinction offered here derives from Heidegger [15]. For Heidegger, meaning cannot be uncovered in the structure of a thing, however complex. The semantic content does not exist in the thing. Meaning-structure is, rather, latent in experience. In other words, meaning emerges from one's interaction with it, emerges alongside our experiences with it:

> *"…meaningful objects…among which we live are not a model of the world stored in our mind or brain; they are the world itself."[9]*

For Heidegger, language was a primal dimension of his ontological pursuit of Being, for words, as translucent bearers of meaning, point to something beyond themselves [16]. Tags, being ontic entities, can only serve as entry points into the complex networks of conceptual associations within our cognition, that is, our ontologies.

4 THE ONTOLOGY OF TAGS

From a Heideggerian perspective tags would be signs. What is the being of signs? "Being-a-sign-for can be formalized as a universal kind of relation, so that the sign-structure itself provides an ontological clue for 'characterizing' any entity whatsoever" (Being and Time, p. 107-108; H. 77). Tags, as signs, are items of equipment whose specific character consists in showing or indicating. Indicating is a referring, but not all referring is indicating. Tags are not references, per se, but rather indicators for the cognitive schemas that are activated upon encountering the tag. When we encounter a tag, as when we encounter a sign, our activated schemas make salient parts of the environment in which it is embedded, and the encounter orients us in a particular way, making us ready to engage 'what is coming.' Tags indicate where one's concern dwells, what sort of involvement one has with something. Tags form entry points into our complex of cognitive and cultural schemas that shape our ontological commitments to the world in which we are immersed.

In terms of creating tags, when we use them for personal recall, we are identifying the salient qualities and dimensions of our experience with the phenomenon or entity being tagged. From the ontological, we create the ontic sign—the tag. They are meaningful to us because they are created based on how we understand the phenomenon, which is in turn based on our personal historical context. Tags become an indicator of that salient experience. They allow us to reactivate our ontological understanding (i.e., activate our schemas) in later encounters with the tags that we create.

We are not only creators of tags, but also consumers of them. The lexical quality of a tag makes it present-at-hand, that which is the focus of our attention—what we are thinking about without all of the background also coming into focus. Tags are indicators of what Heidegger calls ready-to-hand, that which is ready to be used without theorizing about it—the ever-ready emergent evocation of our ontological conceptualizations and commitments. It is the readiness-to-hand quality of tags that evoke the cognitive and cultural schemas that connect us with the tag and to that which it indicates and provides the space where the semantics of a set of tags can be discerned.

If folksonomies are to serve as supplements or complements to formal ontologies, we must be able to disaggregate the sets of tags into cultural identity perspectives, each of which entails the ontological commitments of the culture. But in doing so, we must not mistake the ontic tag representation for the ontological cognitive conceptualization, the extrapersonal lexical structural unit for the intrapersonal schemas it may evoke. Eventually, we want to be able to utilize tag sets in information systems in order to facilitate intercultural understanding, so we must remain aware of the need for interaction, construction and communication mentioned earlier.

REFERENCES

1. Buffa, M. and Gandon, F., SweetWiki: semantic web enabled technologies in Wiki. in *WikiSym'06: Proceedings of the international symposium on Symposium on Wikis*, (2006), ACM, 69-78.
2. Capocci, A. and Caldarelli, G. Folksonomies and clustering in the collaborative system CiteULike, 2007.
3. Cattuto, C., Baldassarri, A., Servedio, V. and Loreto, V. Vocabulary growth in collaborative tagging systems, 2007.
4. Cattuto, C., Loreto, V. and Pietronero, L. Collaborative Tagging and Semiotic Dynamics, 2006.
5. Cattuto, C., Loreto, V. and Pietronero, L. From the Cover: Semiotic dynamics and collaborative tagging. *PNAS*, *104* (5). 1461-1464.
6. Choy, S.-O. and Lui, A.K., Web Information Retrieval in Collaborative Tagging Systems. in *Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on*, (2006), 352-355.
7. Cudré-Mauroux, P., Aberer, K., Abdelmoty, A., Catarci, T., Damiani, E., Illaramendi, A., Jarrar, M., Meersman, R., Neuhold, E., Parent, C., Sattler, K.-U., Scannapieco, M., Spaccapietra, S., Spyns, P. and De Tré, G. Viewpoints on Emergent Semantics. in *Journal on Data Semantics VI*, 2006, 1-27.
8. D'Andrade, R. *The Development of Cognitive Anthropology*. Cambridge University Press, Cambridge, 1995.
9. Dreyfus, H.L. *What Computers Still Can't Do: A Critique of Artificial Reason*. The MIT Press, Cambridge, 1992.
10. Fauconnier, G. and Turner, M. Conceptual Integration Networks. *Cognitive Science*, *22* (2). 133-187.
11. Fisher, K. Locating Frames in the Discursive Universe *Sociological Research Online, 2(3)*, 1997.
12. Furnas, G., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C. and Naaman, M., Why do tagging systems work? in *CHI '06: CHI '06*

extended abstracts on Human factors in computing systems, (2006), ACM, 36-39.
13. Golder, S. and Huberman, B. The Structure of Collaborative Tagging Systems. *Journal of Information Science*, *32* (2). 198-208.
14. Halpin, H., Robu, V. and Shepherd, H., The complex dynamics of collaborative tagging. in *WWW '07: Proceedings of the 16th international conference on World Wide Web*, (2007), ACM Press, 211-220.
15. Heidegger, M. *Being and Time*. Harper and Row, New York, 1927.
16. Heidegger, M. *The Question Concerning Technology and Other Essays*. Harper & Row, Publishers, New York, 1977.
17. Hotho, A., Jäschke, R., Schmitz, C. and Stumme, G., FolkRank: A Ranking Algorithm for Folksonomies. in *Proc. FGIR 2006*, (2006).
18. Hveinden, B. *Divided Against Itself: A Study of Integration in Welfare Bureaucracy*. Scandinavian University Press, Oslo, 1994.
19. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L. and Stumme, G. Tag Recommendations in Folksonomies. in *Knowledge Discovery in Databases: PKDD 2007*, 2007, 506-514.
20. Kipp, M. and Campbell, G. Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices.
21. Lambiotte, R. and Ausloos, M. Collaborative tagging as a tripartite network, 2005.
22. Mattesich, P.W., Murray-Close, M. and Monsey, B. *Collaboration: What Makes It Work*. Amherst H. Wilder Foundation, St. Paul, MN, 2001.
23. Mazzocchi, S. Folksologies: de-idealizing ontologies *Stefano's Linotype*, 2005.
24. Moscovici, S. The Phenomenon of Social Representations. in Farr, R.M. and Moscovici, S. eds. *Social Representations*, Cambridge University Press, London, 1984.
25. Quintarelli, E., Folksonomies: power to the people. in, (2005), Universit√† di Milano.
26. Saab, D.J., Maldonado, E., Orendovici, R., Tchouakeu, L.-M., van Gorp, A., Zhao, K., Maitland, C. and Tapia, A.H., Building global bridges: Coordination bodies for improved information sharing among humanitarian relief agencies. in *5th International ISCRAM Conference*, (Washington, DC, USA, May 4-7, 2008, 2008), 471-483.
27. Santos-Neto, E., Ripeanu, M. and Iamnitchi, A. Tracking User Attention in Collaborative Tagging Communities, 2007.
28. Schmitz, C., Hotho, A., Jäschke, R. and Stumme, G., Mining Association Rules in Folksonomies. in *Data Science and Classification. Proceedings of the 10th IFCS Conf.*, (Heidelberg, 2006), Springer, 261-270.
29. Schmitz, P., Inducing ontology from Flickr tags. in *Proc. of the Collaborative Web Tagging Workshop (WWW ,Äô06)*, (2006).
30. Shirky, C. Ontology is overrated: Categories, links and tags *Clay Shirky's Writings about the Internet, shirky.com*, shirky.com, 2005.
31. Specia, L. and Motta, E. Integrating Folksonomies with the Semantic Web. in *The Semantic Web: Research and Applications*, 2007, 624-639.
32. Sperber, D. and Wilson, D. The mapping between mental and public lexicon. in Carruthers, P. and Boucher, J. eds. *Thought and Language*, Cambridge University Press, Cambridge, 1998.
33. Spyns, P., de Moor, A., Vandenbussche, J. and Meersman, R. From Folksologies to Ontologies: How the Twain Meet. in *On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE*, 2006, 738-755.
34. Staab, S., Santini, S., Nack, F., Steels, L. and Maedche, A. Emergent semantics. *IEEE Intelligent Systems*, *17* (1). 78-86.
35. Strauss, C. and Quinn, N. *A cognitive theory of cultural meaning*. Cambridge University Press, Cambridge, 1997.
36. Talmy, L. The cognitive culture system. *Monist*, *78* (1).
37. Tanaka, J. and Taylor, M. Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder? *Cognitive Psychology*, *23* (3). 457-482.
38. Vander Wal, T. Explaining and Showing Broad and Narrow Folksonomies *Off the Top, vanderwal.net*, vanderwal.net, 2005.
39. Vander Wal, T. Online Information Folksonomy: Presentation Posted *Personal InfoCloud, personalinfocloud.com*, personalinfocloud.com, 2006.
40. Wang, X., Bai, R. and Liao, J. Chinese Weblog Pages Classification Based on Folksonomy and Support Vector Machines. in *Autonomous Intelligent Systems: Multi-Agents and Data Mining*, 2007, 309-321.
41. Wood, D. and Gray, B. Toward a Comprehensive Theory of Collaboration. *Journal of Applied Behavioral Science*, *27* (2). 139-162.

Deconstructing Motivations of ICT Adoption and Use: A Theoretical Model and its Application to Social ICT

Michael J. Scialdone
Syracuse University
School of Information Studies
337 Hinds Hall Syracuse, NY 13244
+13152697283
mjsciald@syr.edu

Ping Zhang
Syracuse University
School of Information Studies
328 Hinds Hall Syracuse, NY 13244
+13154435617
pzhang@syr.edu

ABSTRACT
This paper begins by presenting a case that models of behavioral intention do not provide insight into the core needs of human beings, and as such, cannot inform designers as to what types of ICT features and functions might help meet users' ICT adoption rationale. We review various motivational concepts across different disciplines. We then present a theoretical model of motivation that encapsulates the process from primitive basic human needs, to the formation and attainment of a specific end-state. We use a real life scenario to show the model's explanatory power. We end by discussing potential implementations of this model into the study of ICTs.

Keywords
Motivation, motive, intent, behavioral intention, goals, theory, social ICTs, IT adoption, IT use

1. INTRODUCTION
The labels Google Generation, e-Gen, Facebook Generation, Digital Generation, and Wired Generation represent the ubiquity, pervasiveness, and significance of information communication technology (ICT) use by youth and other populations in their daily lives [12]. Yet, very little is known about the motivations for adoption of ICTs, including social ICTs. Motivation is the energizing force behind behavior intention [1]; it is the conscious or unconscious stimulus for action toward a desired goal, especially as resulting from psychological or social factors [25]. Clear understanding of motivations for adoption is important not only for understanding the e-Gen phenomenon, but also to offer designers of social ICTs practical guidance as to how to construct ICTs that are desirable to use.

The most widely used theoretical models for human behaviors, including ICT adoption, are the theory of reasoned action (TRA) [2, 10] and the theory of planned behavior (TPB) [1], as well as adaptations and variations built from their foundation. As we will demonstrate next, these models fall short in providing the explanatory power behind behavior to offer designers practical advice to build ICTs that individuals are motivated to adopt and use.

TRA posits that behavioral intention is dependent on two factors: subjective norm and attitude toward the behavior [2, 10]. Subjective norm, Ajzen [1] notes, is "the likelihood that important referent individuals or groups approve or disapprove of performing a given behavior" (p. 195). Attitude is a behavioral belief held about a particular object that is linked to a particular outcome [10]. Ajzen [1] observes that a major limitation of TRA is its inability to predict "behaviors over which people have incomplete volitional control" (p. 181). As such, the theory of planned behavior (TPB) is conceived to address the judgments that one has as to how well he or she can engage in actions necessary for an enactment of a behavior. In other words, TPB adds control belief to the model of behavioral intention. Unlike subjective norm and attitude, perceived behavioral control, combined with behavioral intention, is a predictor of behavior itself.

While both TRA and TPB use beliefs to predict the degree of intention to perform a behavior (or what Ajzen [1] refers to as motivation), they fail to capture elements of volitional behavior which are not bound up in beliefs, such as needs and desires. As Gollwitzer [11] notes, "being motivated' implies a number of different phenomena" (p. 53) that begin with desires, and ends with evaluating the achieved action outcome. We will use the following scenario to illustrate the limitations of TRA and TPB.

Joe has recently moved from Italy to the United States to attend college, leaving behind his family, girlfriend and many close friends. It is rather expensive for him to call or send text messages to them regularly. Joe is considering joining the social networking web site Facebook, as many of those he would like to keep in touch with are members. They have often talked about how easy it is to join and use to connect with others. He has, however, had a negative experience with MySpace, another social networking site. He found the website interface clumsy and difficult to navigate, and as such, his interaction with his MySpace contacts was sporadic at best. As such, Joe is dubious about joining Facebook, but sees it as potentially the most viable approach to maintain desired relationships.

Within the scope of TRA and TPB, knowing Joe's beliefs is necessary to predict behavioral intention and behavior in regard to joining Facebook. While his behavioral beliefs are mixed (weighing negative experience against probability of successful interaction with friends and family), his normative beliefs are positive and strong (as he knows many existing members), as are his control beliefs (as he's heard Facebook is easy to join and use). Hence, one could reasonably predict that Joe becomes a member. However, these beliefs do not capture innate needs and personal desires, which are at the core of volitional behavior. In other words, TRA and TBP do not explicitly address the core

needs that may compel specific human activity. As such, an understanding of motivation, as a process that encapsulates needs as they are awoken into desires, ultimately forming intention to reach a goal-driven end state, begs for additional investigation. Such an understanding could offer designers of social ICTs, like Facebook, guidance into implementing designs that motivate individuals to adopt and/or use them.

We now review the literature from a number of disciplines to unearth some fundamental concepts in motivation. Then we present a theoretical model of motivation that traces what we call primitive motives (those internal, fundamental biological and psychological human needs) from their root within individuals, ultimately to the attainment of a specific goal. We illustrate the practical value of the model in offering ICT developers insight as to how design elements might best meet human needs by directing their focus on the amplification of primitive motives to specific (or what we later refer to as objectified) motives, and how intention (in particular, goal intentions, which is also defined below) develops from such motives.

2. OVERVIEW OF MOTIVATION CONCEPTS ACROSS SCHOLARLY LITERATURE

In order to gain a holistic understanding of the various phenomena that have been investigated, captured, and reported when studying motivation, we reviewed the literature in a set of diverse disciplines where human behaviors are of great interest: law, psychology, information systems (IS), human-computer interaction (HCI), and a few other branches of social science. Although we found inconsistencies in terminology use and emphases of investigations within and between disciplines, there is consistency on the existence of some important concepts. These are identified and summarized below.

2.1 Motives

The discipline of law provides some valuable insight regarding the concept of motive. Although this is not without confusion and interchangeable use between the terms motive and intent [3, 17], established distinctions are made by some. Binder [3] notes that following the utilitarianism philosophy which held that motives were purely desiderative states, and that intent was a cognitive state, inconsistent (and thus somewhat interchangeable) usage of the terms began in the 20th century. Arguments were made that a purely cognitive conceptualization of intent was incongruent with ordinary usage and actual legal usage. Such arguments were founded on the idea that intended consequences had to be those that were desired, and thus motive was not distinguishable from intent. Binder [3] explains that motive, in this line of reasoning, was "a kind of intent, one that was more distant or ulterior relative to some more immediate intent" (p. 46).

Such an argument can be observed in the writings of Mercier [23]. He states that there are two parts to the definition of a crime, "the outward act, and the state of mind with accompanies it" (p. 3), and clearly observes that motive is the part of this state of mind, being a series of desires, from primitive instincts to the specific desire for the action. As such, Mercier [23] expresses that motives are more distant intentions, and intentions are motives that are more proximate. The rationale behind this is that desires are interwoven with intent because the consequence of an intended act springs from one's desire to perform the act. This line of reasoning in law suggests that desiderative states can be both primitive (and thus unconscious) and cognitive (and thus conscious).

From a psychological perspective, Gollwitzer [11] positions motivation as a process that involves numerous phenomena, and describes these in a model of phases, known as the Rubicon model [13-15]: action as a temporal, horizontal path that begins with motives, and ends with evaluating the achieved action outcome. He distinguishes these from wishes, implying that motives are basic, fundamental needs, and that wishes are desires that individuals are aware of. Winter et al. [29] explain that psychoanalytic tradition considers motives to be biological, fundamental drives, even though they use other terms such as desires and wishes synonymously. While there may be some degree of inconsistency in terminology, this does illustrate that psychology recognizes raw, unconscious human needs as an element of motivation. For example, Maslow [21], in discussing his hierarchy of needs, explained that primitive, unconscious needs are often at the root of desires that human beings are consciously aware of.

Within the HCI framework of activity theory, Kaptelinin and Nardi [16] state that, "objects of activities are prospective outcomes that motivate and direct activities, around which activities are coordinated, and in which activities are crystallized in a final form when the activities are complete" (p. 66). They further explain that "objects can be physical (such as a bull's eye on a target) or ideal ('I want to become a brain surgeon')" (p. 67). Objects of activities can also be referred to as objectives, giving meaning to what people do. They refer to the work of Leontiev [19], who asserted that needs are biological and/or psychological. As such, an unobjectified need is one that is a primitive state that does not have direction or purpose, while an objectified need is one that has purpose and requires an activity to fulfill.

Clearly, across these three disciplines, there is evidence to suggest that desiderative phenomenon, such as those captured in terms like "needs", "desires", and "wishes" can be categorized into those which are innate and primitive, and therefore unconscious; and those which individuals are aware of, and are about something. In law, Mercier [23], observes that action is the result of both primitive instincts and specific desires. This is reminiscent of the distinction Kaptelinin and Nardi [16] make in HCI between unobjectified needs and objectified needs. Those biological and psychological needs without objects (or goals) are those without direction, leading us to the conclusion that those which are objectified lead to intention. Likewise, in psychology, Gollwitzer [11] makes the distinction between motives and wishes, the former being desires that one is not aware of, and the latter being those that one is aware of.

Considering this point of general consistency, we have chosen to conceptualize two different types of motives in our model to maintain this distinction. We define the term *primitive motive* in our model as those unconscious needs that are most fundamental to human existence, which may be either biological or psychological. O*bjectified motive*, we define then, as consciously recognized desires (stemming from primitive motives) toward achievement (or avoidance) of a particular end. We adapt this

term from activity theory [5] to refer to this need which has been given direction. As Nardi [24] recognized that objects, as goals, are fundamentally rooted in context, we similarly see that it is context that amplifies primitive motives into objectified motives.

Similar to how Gollwitzer's [11] describes wishes leading to goal intention, our model depicts an objectified motive leading to the same, consistent with the spirit of activity theory depicting an object giving direction toward a particular end [16]. As such, we now turn our attention to defining what intention is.

2.2 Intentions

In criminal law, Binder [3] provides a historical overview on how motive and intent came to be relevant to the judicial system. Motive first came to be distinguished from intent in the late 18th century because courts were urged to distinguish between character and behavior. Motive was originally associated with character, while "intentions" were associated with behavior, which could then be compared to written rules of conduct. The utilitarian school of thought in the mid 19th century perceived motives as desiderative states, or as Binder [3] explains, "a desire or fear that causes action" (p.31), while intentions were described as cognitive states, or "expectations that accompany action" (p.31). Chiu [4], who equates motives with desires, observes that an act without a motive is one which is unintentional. Hence, under these arguments made by these law scholars, we can see that there is a distinction between two concepts. Of note, as the scholarly literature in law reviewed here seems to use the terms intent and intention interchangeably, we favor the term intention as it is more consistent with usage in other disciplines.

The most widely-adopted and expanded model of usage in IS, the technology acceptance model (TAM) [7-9], is derived from TRA/TPB. As such, usage of these terms in both psychology and IS tends to be consistent. In writing about TPB, Ajzen [1] states that "intentions are assumed to capture the motivational factors that influence a behavior; they are indications of how hard people are willing to try, of how much of an effort they are planning to exert, in order to perform the behavior," (p. 181). In Ajzen's conceptualization, there is only one type of intention, behavioral intention, as behavior is what intention is directed toward.

However, Gollwitzer [11] points out a different type of intention, that of goal intention. He explains that even with high desirability and feasibility, an individual needs to form determination in order to turn a wish (or what we call an objectified motive) into an intention. As he defines it, *goal intention* is the sense of obligation that an individual forms in order to reach the particular desired ends specified by the wish. Goal intention has similarly been called intent in law. For example, Mercier [23] and Kugler [18] spoke of intent as directed toward a particular goal. Behavioral intention, Gollwitzer [11] explains, proceeds goal intention, which he distinguishes as focusing on the behavior requisite for goal perusal. This is similar to the usage of intention that Ajzen [1] employs in that it is directed toward behavior. Thus, we find it prudent to define *behavioral intention* as commitment to a specific implementation course toward volitional goal achievement that includes how, when, and where to act, as well as limitations on duration and effort.

2.3 Goals

As goal intention promotes the formation of a goal, and produces behavioral intention, it is important to address what a goal is. Much as the law literature inconsistently and interchangeably uses the terms motive and intent, the term goal is not well specified. Binder [3] often refers to intentions as goals, while Mercier [23] refers to goals as the termination of a purposeful action. Kugler [18], meanwhile, refers to goals as the ultimate aim of motive. In the two latter cases, a goal is conceived of as some type of end state, either as the termination of an action, or as an ultimate aim. This fits in closely with the usage of goals in psychology and HCI.

Within psychology, Maslow [21] writes that goals are inseparable from motives, observing that goals are the ends toward which motives drive intention. From an HCI perspective, Nardi [24] notes that, "the word goal in everyday English usage is generally something like what activity theorists call an object in that it connotes a higher-level motive" (p. 48). As Christiansen [5] coined the term "objectified motive" as a way of denoting purpose, an object (as in objective) in this reasoning represents the goal, while an objectified motive is the driver of action toward this goal. This is similar to the psychology perspective of Locke and Latham [20] who define goal as the "object or aim of an action" (p. 175), serving to direct and energize actions and enforce persistence. In both disciplines, goals are essentially perceived as end states. As motive (specifically objectified motives) direct one's attention toward goal intention, the goals that individuals strive to attain are therefore desired. As such, we define a *goal* as a desired end state.

2.4 Behavior

While goal intentions form goals, as noted above, they also form behavioral intentions. Unlike other terms reviewed in this paper, the conceptualization and usage of behavior is essentially consistent throughout the disciplines reviewed in this paper.

Writing from a psychological perspective, Coon [6] broadly defines behavior as anything that humans do, from sleeping, talking, sneezing, or thinking. He describes them as both activities and actions. In law, behavior is, in fact, the entire foundation of the discipline, as the purpose of law is to regulate (or guide) everyday behavior [22, 27]. As acts and actions are referred to as the outward, observable factors in law that determine culpability [3, 23], they are the core of behavior. Activity theory, which approaches behavior from an HCI perspective, sees behavior as the performance of activities through actions [24]. What these latter two perspectives capture, which the first does not, is behavior which is under an individual's own control. However, Wehmeyer [28], who has a psychological perspective on behavior, does use the term volitional for those behaviors which are purposeful "acts that enable the actor to cause things to happen in his or her life" (p. 115).

As those behaviors which are meant for inclusion in our model are premeditated through behavioral intention, we assume that they are volitional and thus under the control of individuals. We also assume that as a goal intention has already been formed, that they are directed toward, and performed in expectation of, attainment of a particular goal. As such, we choose define *behavior* within our model as volitional actions directed toward goal attainment.

2.5 Summary of Literature Reviewed

Our review of the literature across law, psychology, and IS/HCI has led us to the distinction and definition of two types of motives and two types of intentions, as well as a clear conceptualization of behaviors and goals. As the concepts we identified are influenced from motives, through intentions, to the performance of behavior, and ultimately goal attainment, a process emerges. Thus, we adopt the term *motivation*, to signify this process that encapsulates motives, intentions, behaviors, and goals.

3. THE MOTIVATION MODEL

Consistent with the scholarly literature above, we consider motivation to be a process. Our model is designed to depict the internal process mechanisms that lead to observable, motivated behaviors. By deconstructing motivation into such a detailed process, we hope to identify elements that can lead to informed ICT design decisions. Figure 1 illustrates our model, while an explanation of the model follows.

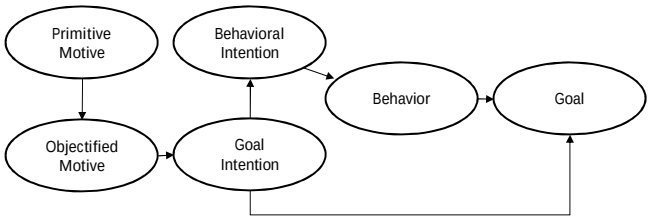


Figure 1. A Motivation Model of ICT Adoption and Use

Our model begins with *primitive motives*, conceived of as biological or psychological needs that are fundamental to existence. These are context-free desires, common and inherently preprogrammed into the human organism. *Objectified motives* are consciously recognized specific desires which direct attention toward a particular goal.

Primitive motives are transformed into objectified motives through the influence of contextual factors that make an individual aware of a need. As desires that are consciously recognized, if an objective motive has high enough desirability and feasibility, a *goal intention* is formed. This is the sense of obligation an individual creates for him or herself toward reaching that desire specified by the objectified motive. A goal intention specifies a *goal*, an end state which represents the fulfillment of the specific desire, and leads to *behavioral intention*. This is defined as commitment to a specific implementation course toward the desired goal, consisting of when, where, and how to act, including limitations on effort and duration. The behavioral intention directs the *behavior* that the individual engages in, which is conceived of as those volitional actions which are performed toward attainment of the specified goal.

4. APPLYING THE MOTIVATION MODEL

To illustrate how our model may have implications for ICT design and use, we turn to the scenario introduced earlier in order to identify the various constructs and their relationships.

Per our scenario, Joe has the objectified motive of wishing to keep in touch with those he left behind in Italy. Yet, unlike existing models that predict or explain behavior, such as TRA and TBP [1, 2, 10], our model allows us to look back one step further. In other words, what primitive motives give form to this objectified motive? One which is evident is that of relatedness [26, 30], a common, basic need to maintain social connections. With the context of Joe being away from his family and friends in Italy, his primitive motive of relatedness becomes amplified into his objectified motive, a conscious wish to keep in touch with them. Of note, we do recognize that this could be looked at as multiple (albeit similar) objectified motives, such as Joe wishing to keep in touch with family, Joe wishing to maintain a relationship with his girlfriend, and Joe wishing to continue to interact with friends on some level. For the sake of parsimony, and as all three of these objectified motives can fit into the attainment of a single goal, we choose to refer to it as a single motive.

An objectified motive can be turned into a goal intention if desirability is strong enough, and if it is feasible enough. In Joe's case, despite having reservations about joining another social ICT website, if his desirability to keep in touch with family and friends is strong enough, a goal intention will form. This goal intention will be a sense of obligation for Joe to join Facebook, while also making him aware that joining Facebook is a goal he must work toward. It is this sense of recognition, of having an unattained goal, which spawns Joe's behavioral intention to join, consisting of a particular course of action that involves the what, where, and how, as well as the effort he will put into it. For example, his behavioral intention to join Facebook may consist of logging into his computer at home tomorrow evening after class, and filling out the requisite online forms. Of note, behavioral intention does not necessarily mean that the behavior will occur. Joe may accidentally fall down stairs on his way home and break his leg, delaying his behavior. He may get a call from his girlfriend breaking-up with him, which may cause him to abandon his plan. Or there may be a power outage in the dorm that causes Joe to adapt his strategy and join from another computer at another time. In this way, behavioral intention is a plan that may not necessarily comes to fruition, but does serve to further direct attention toward a goal.

The behavior Joe performs may or may not match up exactly with that expected within his behavioral intention, based on unexpected influences noted above. However, going forward in the process of motivation, Joe engages in behavior that allows him to reach his goal. The enactment of behavior that is geared toward the goal is a result of behavioral intention, even if outside forces have intervened. As our example above stated, if Joe has intention to join Facebook from his door room at a specified time, but his power goes out then, the behavior he performs by logging into his iPhone and joining Facebook from there, is still a product of his original behavioral intention, even though the actual behavior is slightly different. The process of motivation still continues unbroken in our model. However, if Joe's girlfriend breaks up with him, and causes him to abandon his intention, we would say that the process of motivation has ended. If the goal is attained, we would say that the process of motivation, in regard to this particular objective, has allowed Joe to reach his desired end state. Looking backward we can see that this end state originated from

Joe's primitive motives, and was the energizing force that carried him through to the attainment of his goal.

Our motivational model is, in this way, stronger than that of TRA and TPB in that designers of ICTs recognize that goal attainment (for example adoption or use) is a product of fixed, and common primitive needs. TRA and TPB fail to consider such human biological and psychological needs as drivers toward behavior. As primitive motives are amplified into objectified motives within a particular context or by particular activity (as Nardi [24] sees context and objects as mutually influencing one another), designers of social ICTs might look at how to design contexts that invoke this amplification. For example, if Facebook was to include a romantic, match-making feature that the designers wanted users to adopt, they might subtly put single users in a context that makes them aware of their needs. Designers may tweak the interface to bold users' relationship statuses of "single", or even write a script that embeds pictures of singles that are using the app onto the side of other singles' windows. Similarly, to promote a higher degree of usage, designers may want to add new communication features such as video chat or the ability to exchange files, as these might cause users to feel more socially connected. In other words, by creating contexts that makes one aware of wishes to fulfill psychological or biological needs, adoption and use might be increased.

5. CONCLUSION

As Hardey [12] noted that adoption of social ICTs is a popular enough phenomenon to garner a number of different generational labels for its constituents, understanding the motivational factors involved is necessary to not only explain why, but also to give practical guidance to designers who wish to maximize adoption and use for a given social purpose. In this paper, we reviewed literature that spans disciplines such as law, information systems, psychology, and human-computer interaction to formulate a holistic understanding of what motivation entails. We then developed a model that depicts various, distinct stages of the motivation process. Such deconstruction of motivation allows designers to emphasize certain elements of the model to inform design of desirable ICTs. We used a social ICT (Facebook) to illustrate the main concepts and relationships bound within our model. However, we want to emphasize that it can be applied to other types of ICTs.

There are many potential future research directions. Our model is the first step toward filling the gap in existing theoretical work to explain human behaviors in the ICT context. Certainly, there is a need for further refinement and empirical validation. The formation of objectified motives, goal intention, and behavioral intentions within given contexts are worth further investigation in order for the model to be of value to designers. Additionally, identifying how specific features and affordances of social ICTs allow adopters to meet their goals, and thus human needs, is potentially another very influential direction for research that may ultimately influence ICT design.

6. REFERENCES

[1] Ajzen, I. 1991. The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50. 179-211.

[2] Ajzen, I. and Fishbein, M. 1980. *Understanding Attitudes and Predicting Social Behavior*. Prentice-Hall, Englewood Cliffs, NJ.

[3] Binder, G. 2003. The Rhetoric of Motive and Intent. *Buffalo Criminal Law Review*, 6 (1). 1-96.

[4] Chiu, E.M. 2005. The Challenge of Motive in the Criminal Law. *Legal Studies Research Paper Series* (February 2005).

[5] Christiansen, E. 1996. Tamed By a Rose: Computers in Human Activity. in Nardi, B. ed. *Context and Consciousness: Activity Theory and Human Computer Interaction*, The MIT Press, Cambridge, MA.

[6] Coon, D. 1994. *Essentials of Psychology: Exploration and Application*. West Publishing Company, St. Paul, MN.

[7] Davis, F.D. 1986. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. *Sloan School of Management*, Massachusetts Institute of Technology.

[8] Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35 (8). 982-1003.

[9] Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. 1992. Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. *Journal of Applied Social Psychology*, 22 (14). 1111-1132.

[10] Fishbein, M. and Ajzen, I. 1975. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA.

[11] Gollwitzer, P.M. 1990. Action Phases and Mind Sets. in Higgins, E.T. and Sorrentino, R.M. eds. *Handbook of Motivation and Cognition: Foundationd of Social Behavior*, Guilford Press, New York.

[12] Hardey, M. 2009. ICTs and Generations - Constantly Connected Social Lives. in *COST: The Good, the Bad and the Challenging*, (Copenhagen, Denmark, 2009).

[13] Heckhausen, H. 1987. Wishing-Weighing-Willing. in Heckhausen, H., Gollwitzer, P.M. and Weinert, F.E. eds. Jenseits *des Rubikon: Der Wille in den Hunzantaissenschaften*. Springer-Verlag., Heidelberg, 3-9.

[14] Heckhausen, H. and Gollwitzer, P.M. 1986. Information Processing Before and After the Formation of an Intent. in Klix, F. and Hagendorf, H. eds. *In memoriam Hermann Ebbinghaus: Symposium on the structure and function of human memory*, Elsevier/North-Holland, Amsterdam, 1071-1082.

[15] Heckhausen, H. and Gollwitzer, P.M. 1987. Thought Contents and Cognitive Functioning in Motivation versus volitional States of Mind. *Motivation and Emotion* (11). 101-120.

[16] Kaptelinin, V. and Nardi, B.A. 2006. *Acting with Technology: Activity Theory and Interaction Design*. The MIT Press, Cambridge.

[17] Kaufman, W.R.P. 2003. Motive, Intention, and Morality in the Criminal Law. *Criminal Justice Review*, 28 (2). 317-335.

[18] Kugler, I. 2002. *Direct and Oblique Intention in the Criminal Law*. Ashgate Publishing Company, Burlington, VT.

[19] Leontiev, A. 1978. *Activity, Consciousness, and Personality*. Prentice-Hall, Englewood Cliffs, NJ.

[20] Locke, E.A. and Latham, G.P. 2002. Building a Practically Useful Theory of Goal Setting and Task Motivation. *American Psychologist*, 57 (9). 705-717.

[21] Maslow, A.H. 1943. Preface to Motivation Theory. Psychosomatic *Medicine*, 5. 85-92.

[22] Melton, G.B. 1992. The Law is a Good Thing (Psychology Is, Too): Human Rights in Psychological Jurisprudence. *Law and Human Behavior*, 16 (4). 381-398.

[23] Mercier, C. 1926. *Criminal Responsibility*. Physicians and Surgeons Book Co., New York.

[24] Nardi, B.A. 1992. Studying context: a comparison of activity theory, situation action models, and distributed cognition *East-West HCI Conference*, ICSTI, Moscow.

[25] Oxford English Dictionary (2009), Motivation. Retrieved November 10, 2009 from http://dictionary.oed.com/cgi/entry/00316490?single=1&query_type=word&queryword=motivation&first=1&max_to_show=10

[26] Reeve, J. 2001. *Understanding Motivation and Emotion*. John Wiley & Sons, Inc., University of Iowa.

[27] Teubner, G. 1983. Substantive and Reflexive Elements in Modern Law. *Law & Society Review*, 17 (2). 239-285.

[28] Wehmeyer, M.L. 2005. Self-Determination and Individuals with Severe Disabilities: Re-examining Meanings and Misinterpretations. *Research & Practice for Persons with Severe Disabilities*, 30 (3). 113-120.

[29] Winter, D.G., Stewart, A.J., John, O.P., Klohnen, E.C. and Duncan, L.E. 1998. Traits and Motives: Toward an Integration of Two Traditions in Personality Research. *Psychological Review*, 105 (2). 230-250.

[30] Zhang, P. 2008. Toward a Positive Design Theory: Principles for Designing Motivating Information and Communication Technology. in Avital, M., Bolland, R. and Cooperrider, D. eds. *Designing Information and Organizations with a Positive Lens, a volume of the Advances in Appreciative Inquiry series*, Elsevier.

Motivated Information Behavior

David Schwieder
Graduate School of Library and Information Science
501 E. Daniel, Champaign, IL 61820
schwied1@illinois.edu

ABSTRACT

While motivation is recognized as central to various aspects of information behavior, motives remain surprisingly underemphasized in information behavior research. Major theories focus almost exclusively on other psychological elements, primarily cognition, while studies of motivation have been limited or absent in a variety of important respects. In this paper, I suggest that a stronger emphasis on motives is warranted. Drawing on recent trends in social psychology research, I argue that a "motivated information behavior" approach can offer a variety of benefits: it can improve our explanations of information behavior, unify disparate research areas, and illuminate some of the mechanisms underlying important information behavior phenomena.

Keywords

motivation, information behavior, information needs, information seeking, information avoidance

INTRODUCTION

Motivation plays a central part in information behavior. Information needs are commonly assumed to rest on a motivational base, and many formulations of this needs concept invoke a motivational drive. These motives then spark information seeking; as Donald Case notes [1] "information seeking is a catchall phrase that encompasses a variety of behaviors seemingly motivated by the recognition of "missing" information." Similarly, much of the work on the related topic of information avoidance recognizes the important role that motives play in this process.

Despite this acknowledged importance, however, motivation remains underdeveloped in LIS information behavior research. Major theories pay little attention to motives, focusing instead on other psychological factors, primarily cognition. When motivation is examined, it is often vaguely conceptualized, or analyzed through scattered, ad hoc treatments which focus on specific motives relevant to particular areas. Unfortunately, these approaches have failed to identify more fundamental aspects of motivation or to capture its core importance in information behavior.

In this paper, I argue that existing information behavior work contains the seeds of a more productive approach. Drawing on recent social psychology studies, I suggest that a focus on "motivated information behavior" (MIB) can move LIS research in a more productive direction. While simple in itself, MIB offers a variety of important advances; better explanation, a framework for organizing disparate research areas, useful new predictions, and the illumination of some key mechanisms underlying important information behaviors.

The (Relative) Neglect of Motivation

Information Seeking

Motivation has never played a particularly prominent part in information seeking research. Along with other psychological factors, motives were largely ignored by the "system-centered" approaches that dominated early information retrieval study. These early approaches emphasized the rational, purposive dissemination of information"typically treating users as little more than passive receivers"and thus psychological factors were seen as largely irrelevant [2].

Nor did the "user-centered" turn in the 1970s and 1980s introduce an emphasis on motives. This is most apparent in the major theories that have gained prominence in the information behavior field. To one degree or another, most of these theories place a central emphasis on cognition. Nick Belkin's ASK formulation was explicitly cognitive, with other theories like Robert S. Taylor's model and Brenda Dervin's sense-making approach also centered primarily around cognitive concerns [3,4,5]; reviewed in [6,7]. Another approach, Carol Kuhlthau's Information Search Process model, (ISP), incorporates affect while retaining a strong focus on cognition [8].

The lack of theoretical focus on motivation is readily apparent in a recent overview of information behavior theory. Of the seventy-some theories outlined in Theories of Information Behavior, only one explicitly invokes "motivation" in the title. Moreover, while a handful of these

theories concern subjects with a clear connection to motivation, like monitoring and blunting, and library anxiety, the presentations of these theories make no explicit references to this term. Nor is motivation mentioned in any the thirteen "meta-theories" discussed in the opening chapter of the book [9].

Of course, motives have not been entirely absent from information behavior research. The most popular approach has been to identify specific motives thought relevant for particular informational contexts or tasks. Studies of Web-based activities, for example, have examined motives like entertainment and communication that are seen to drive Web use [10,11], while other studies have analyzed motives in educational settings [12,13]. The number of cited motives varies widely; some studies employ a general, unspecified information seeking drive, while others identify dozens of specific motives [14,15].[1] These studies have been useful, but they have neither yielded broader theory nor identified what one can call "basic motives more fundamental to the nature of human desire than particular motives that are the result of relatively specific conditions [16]."

Information Avoidance

If information seeking research has paid little explicit attention to motives, motivational considerations have been more prominent in studies of information avoidance. Situations where people desire to avoid threatening or ego-challenging information seem to prompt an intuitive recognition that motivation plays an important part, and a number of studies have analyzed the dynamics of this process [17]. Many of these studies have focused on the areas of medicine and health, where "bad news" can have dire consequences.

However this research has suffered from several problems in terms of motivation. Motives have rarely been well specified, and many studies simply posit a vague drive to avoid information. Little attention has been paid to theory. More fundamentally, information avoidance studies have been a minor sidelight to information seeking work; as Donald Case notes, information behavior research has generally rested on the assumption that information seeking represents the norm [18].

[1] If inclusion of motivation has been so sporadic, how do these various theories move from mental factors to action? Some approaches simply proceed as if behavior follows directly from mental activity, while others have suggested that emotions drive behavior. While these approaches are not unreasonable, especially the latter, by definition neither can capture the nature or importance of motivation.

Causes and Consequences

The previous sections document the limited and fragmented attention paid to motivation in the information behavior literature. From a psychological perspective, this is somewhat surprising. Psychology research has traditionally posited a "triumvirate" of three basic mental elements; cognition; affect and emotion; and a third area which includes the interrelated concepts of conation and motivation [19]; from this perspective, the absence of motivation is apparent. This raises an obvious question; why have motives not played a larger part in information behavior research?

The most obvious answer is that motives have been overshadowed by cognition. In part, this seems to be a legacy of the early systems-centered approach. As Savolainen notes, in that type of model "the system is the essential order, and the individual or user bends to it [20]" If the purpose of information systems is to provide a rational supply of information, then the role of the user is to rationally retrieve it; rational systems were seen to beget rational "task oriented" users. And rational users are cognitive users. Reflecting this, psychological approaches have tended to place a strong focus on cognition.[2]

More broadly, cognitive approaches have dominated related disciplines. In the1970s and 1980s, when the user-centered turn was developing in information behavior research, the social psychology subfield was in the midst of an intellectual cycle which strongly asserted the primacy of cognition [21]. This "cognitive imperialism" drew attention away from other factors like affect and motivation. In the case of motivation, the effect was particularly dramatic. A variety of psychological phenomena can be explained in either cognitive or motivational terms, and thus motivational approaches languished behind more favored cognitive rivals [22]. These trends affected a number of disciplines that "import" from social psychology, information behavior research among them.

This under-emphasis on motivation has had several unfortunate consequences. As indicated earlier, the theoretical underspecification of motives has been accompanied by empirical work which is fragmented across a variety of specific motives and particular areas of application. More importantly, these theoretical and empirical limitations have obscured more basic forms of motivation, and their importance in information behavior, as the next section will suggest.

[2] This cognitive focus may have been further advanced by this early work's focus on scientist and engineers, two disciplines which place a strong professional emphasis on rationality and objectivity.

Motivated Information Behavior

While information behavior studies have paid relatively little attention to more fundamental motivations, this does not mean that such motives have been entirely absent from this work; rather, most of these studies have implicitly assumed a basic model of motivation. This can be appreciated through the lens of recent social psychology work. Following several decades where motives were out of fashion, this field has returned to an earlier interest in motivation; as the most recent Handbook of Social Psychology has it, "motivation is back [23]."

The central thrust of this recent work has focused on the motivated construction of understanding. This approach posits two basic types of motives. The first emphasizes accuracy. In many cases, people wish to form accurate beliefs and impressions about the world; that is, to reach conclusions that are best supported by available evidence and information. Given the obvious importance of accuracy in many instances of "sense making," much of this recent psychology work has analyzed accuracy-seeking motives and how they shape people"s thinking and information use [24].

However accuracy is not the only possible goal, and at other times people seek to construct more comforting conceptions of reality [25]. Motives serving these kinds of goals are characterized as "directional," and they help people to arrive at conclusions which they wish to reach. These conclusions can be used to ward off threatening implications, or to defend basic values or existing points of view. Directional motives function primarily through biased strategies for selecting and evaluating beliefs and information; for example, by affecting which types of information are considered or how this information is used [26].

What does this tell us about information behavior? If information seeking and information avoidance are motive-driven, as Donald Case argues, and if these motives will be either accuracy-seeking or directional, as the social psychology research suggests, then many information behaviors will be driven, at a fundamental level, by one or the other of these basic motives.

While information behavior research has not explicitly recognized or labeled these directional or accuracy-seeking motives, in many cases this work has assumed their existence. This is most evident in the information avoidance literature. As indicated earlier, it seems intuitively apparent that a desire to avoid certain types of information is motivationally driven, and information avoidance research has regularly, if rather casually, treated motivation as important. Though the term "directional" is not used, the motives involved here-for example with serious medical conditions-obviously are intended to help people attain or preserve a particular perspective or conclusion; i.e. one that is uninformed by troubling information.

On the other hand, assumptions about accuracy seeking motives have been less apparent. Accuracy motives are not explicitly invoked in the information behavior literature-indeed, seeking accuracy hardly seems like a motive at all-but most information seeking studies have implicitly assumed that people wish to reach accurate conclusions. This assumption comes easily; absent any reason to think otherwise, accuracy seeking constitutes a reasonable, almost automatic default. Moreover, along with being an often functional strategy, seeking accuracy also is a core value in the LIS professional culture; bibliographic systems, reference and instructional services are designed to emphasize accuracy. And if the system prizes accuracy, then, in an odd echo of the early information retrieval literature, it is rather easy to assume that users do too. Accordingly, the belief that people desire accuracy has served as an unrecognized assumption in most research on information seeking and use.[3]

Discussion

What does it buy us to recognize that information behavior is motivated in this basic sense? Most directly, positing "motivated information behavior" can improve theory and explanation. We can appreciate this by considering one of the leading information seeking theories, the ISP model. This model emphasizes "affective (feelings) [and] cognitive (thoughts)," and these two factors interact to shape information seeking behavior [28]. As noted, this approach differs from most theories by its inclusion of affect. But what does this add? If one can model information behavior in purely cognitive terms-as most theories do-what do we gain by adding emotion? The payoff, of course, is richer understanding and explanation. Kuhlthau's research subjects report that feelings are an important aspect of information seeking, and this points to useful insights; how anxiety can short-circuit information seeking, for example, or the discovery that uncertainty actually rises at certain points in the information seeking process rather than simply declining monotonically.

Similarly, then, considering motivation can also contribute additional insights. While information behaviors

[3]The point that unrecognized assumptions guide research is a generally acknowledged one, and accuracy-seeking is not the only implicit assumption in information behavior work. Commenting on the relative lack of attention to information avoidance, Donald Case notes the underlying assumption that people seek information; "As in Aristotle"s time, it is assumed that people want to know; looking for information is a natural aspect of being human" [27].

can be explained without it, motivation can add explanatory and theoretical insights that other psychological factors cannot, particularly when information behavior is viewed as an active process of construction.

Most broadly, motivated information behavior can provide a powerful organizing framework, one that helps to connect disparate literatures and findings. While information seeking and information avoidance have been treated as distinct phenomena, the MIB perspective views these two behaviors as different sides of the same motivational coin, with the choice between them driven by the particular goal that people happen to employ. When they wish to gain an accurate understanding of some unfamiliar area, people will typically tend to seek information; alternatively, they may avoid or reject information when their motives direct them toward reaching or preserving particular positions or states of mind.

The MIB framework can also organize more specific behaviors. While information avoidance has been the only form of "non-use" to receive much LIS attention, MIB suggests that avoidance is hardly unique; rather, it is simply one type of directionally motivated information behavior.[4] Another such behavior involves information rejection. Studies in a number of other fields have examined the use and evaluation of information, and they have consistently found that people show a strong bias toward accepting information consistent with their existing views while rejecting information which clashes with these views or undermines them [30,31,32,33]. Accordingly, a MIB approach can offer predictions about the existence and basic nature of directionally-driven information behaviors.

Following from this, the MIB approach can also help to identify specific mechanisms underlying particular information behaviors. For example, psychology studies suggest that a main mechanism underlying information rejection is "counterarguing," an active endeavor where unpalatable information is challenged or contested, then dismissed through the mobilization and use of reasons or arguments which undermine or dispute undesirable evidence. While people accept affirming information rather uncritically, discordant information is subjected to stringent evaluation [34]. Alternatively, people may simply pay more attention to confirming information than disconfirming information [35]. Along with information avoidance and rejection, MIB can also shed light on information seeking; studies show that people driven by accuracy motives seek information in a more balanced manner, and are resistant to cognitive biases that can undermine effective information seeking.[5]

Turning from theory to practice, does the MIB perspective help to inform practical library work? Yes, but not in the usual sense. While we typically expect research to guide practice, a MIB perspective probably raises more practical questions than it answers. This is due to the prevalence of directional motives. While LIS studies of information avoidance have focused heavily on health concerns, studies in social psychology and other disciplines predict that directional motives-and the behaviors they inspire-will be common, even the norm, in areas where people wish to preserve existing opinions, defend cherished interests or values, or to ward off perceived threats. Clearly, many types of information offered in the library-involving politics, for example, controversial issues and events, or any matters where people are invested in some settled view of the world-can involve the kinds of concerns that tend to prompt directional motivations. These points thus combine to suggest that library-related directional motives and behaviors will be far more common than the "accuracy-assuming" LIS literature has supposed.

This poses an obvious problem; when people reject or avoid information, information literacy will typically suffer. This problem is hardly a novel one, of course; information literacy routinely is compromised by a host of factors, including users' lack of effort, their cognitive biases or anxieties, or limitations in information systems [38, 39, 40, 41]. But motives are different. Whatever the unfortunate effects of these other factors, we can assume that users still have some basic desire to utilize the carefully assembled "public knowledge" that libraries exist to supply. With directional motives, however, we cannot.

The implications are important. While problems stemming from user biases or "least effort" can be addressed through traditional library approaches like instruction, reference services, and improvements that facilitate easier use of information systems, it is not clear how libraries can deal with the rejection or outright avoidance of information. Aside from the practical difficulty of providing services-many people who wish to avoid information will probably just avoid the library-there is a deeper problem as well. Helping anxious or casual users to find relevant information falls squarely within the traditional library paradigm, but serving people who have some basic wish to remain ignorant is another matter. Users who prefer "illusion," to use the social psychology term, would seem

[4] The term "non-use" is from Wilson [29].

[5] Finally, besides contributing to LIS study, applying motivated information behavior theory to this area could also help to extend the theory itself. Given its reliance on experimental methods, social psychology research has tended to focus primarily on information evaluation; experiments present researcher-collected information to subjects, and then analyze its effects. However LIS research can direct attention to the prior information seeking stage, which has received little attention in psychology motivational research. Extending the theory in this manner would help to "export" LIS findings to other disciplines, a useful excercise which has tended to lag in the past [36,37].

to present a unique challenge.

Conclusion

Motivation occupies a paradoxical place in the information behavior literature. Recognized as central for information needs, motives have nonetheless played only a sporadic and fragmented role in research on information seeking and use: it is as if motives provide a powerful initial spur to the process, then virtually disappear. Clearly this is account is unconvincing, and we would expect motivation to influence subsequent stages in the information process.

As indicated, motives can make a significant contribution to the study of information behavior. While simple in itself, the notion of motivated information behavior offers several important benefits: improved explanation, frameworks to organize multiple research areas and findings, and the ability to offer new predictions and illuminate mechanisms of action. This array of benefits is broad, but probably not surprising. If such basic motivation is central, as the social psychology literatures suggest, and if it has remained underemphasized in information behavior work, then incorporating it would be expected to provide a variety of useful dividends.

Obviously this does not mean that motivated information behavior is the only way to conceptualize motives, or that motives must be incorporated into every theory or approach. However it does suggest that many information behavior literatures could profitably include these and other sophisticated treatments of basic motives and motivation.

REFERENCES

[1] D. O. Case. *Looking for information: A survey of research on information seeking, needs and behavior.* Academic Press, New York, 2002.

[2] B. Dervin and M. Nilan. Information needs and uses. In M. Williams (Ed.), *Annual Review of Information Science and Technology*, 21: 1-25, Knowledge Industry,White Plains, NY, 1986.

[3] N.J. Belkin, R.N. Oddy, and H.M. Brooks. ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38: 61-71, 1982.

[4] B. Dervin. *Sense-making methodology reader: Selected writings of Brenda Dervin.* Hampton Press, 2003.

[5] R.S.Taylor. Question-negotiation and information seeking in libraries. *College & Research Libraries*, 178-194, 1968.

[6] C.C. Kuhlthau. Inside the search process: Information seeking from the user"s perspective. *Journal of the American Society for Information Science*, 42, 361-371, 1991.

[7] R. Savolainen. The sense making theory: Reviewing the interests of a user-centered approach to information seeking and use. *Information Processing & Management*, 29, 13-28, 1993.

[8] Kuhlthau, *Inside.*

[9] K.E. Fisher, S. Erdelez, and L. McKechnie. *Theories of Information Behavio*r. Information Today, 2005.

[10] S. Sun, A.M. Rubin, and P.M. Haridakis. The role of motivation and media involvement in explaining internet dependency. *Journal of Broadcasting & Electronic Media*, 2008.

[11] D. Yoon, F. Cropp, and G. Cameron. Building relationships with portal users: The interplay of motivation and relational factors. *Journal of Interactive Advertising*, 3, 2002.

[12] R.V. Small, N. Zakaria, and H. El-Figuigu. Motivational aspects of information literacy skills instruction in community college libraries. *College & Research Libraries*, 65, 96-121, 2004.

[13] J. Heinstrom. Fast surfing for availability or deep diving into quality: motivation and information seeking among middle and high school students. *Information Research*, 11, 2006.

[14] M.J. Dutta-Bergman. The impact of completeness and Web use motivation on the credibility of e-health information. *Journal of Communication*, 253-269, 2004.

[15] T.F. Stafford, M.R. Stafford, and L.L. Shkade. Determining uses and gratifications for the Internet. *Decision Sciences*, 35, 259-288, 2004.

[16] T.S. Pittman. Motivation. In D. Gilbert, S.T. Fiske, & G. Lindzey, (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, 549-590). Oxford University Press, Inc, New York, 1998.

[17] D.O. Case, J.E. Andrews, J.D. Johnson, and S.L. Allard. Avoiding versus seeking: The relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. *J Med Libr Assoc,* 93, 353-362, 2005.

[18] Case, *Avoiding.*

[19] B. Parkinson, and A. Colman. *Emotion and motivation.* Longman, Upper Saddle River, NJ, 1995.

[20] Savolainen, *Reviewing.*

[21] E.E. Jones. Major developments in five decades of social psychology. In D. Gilbert, S.T. Fiske, & G. Lindzey, (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, 3-57). Oxford University Press, Inc, New York, 1998.

[22] Z. Kunda. The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498, 1990.

[23] Pittman, *Motivation.*

[24] Pittman, *Motivation.*

[25] Pittman, *Motivation.*

[26] Kunda, *Case.*

[27] Case, *Avoiding.*

[28] Kuhlthau, *Inside.*

[29] P. Wilson, P. Unused relevant information in research and development. *Journal of the American Society for Information Science*, 46, 45-51, 1995.

[30] C.G. Lord, L.Ross, and M.R. Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098-2109, 1979.

[31] C.S. Taber, and M. Lodge. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50, 755-769, 2006.

[32] R.P. Vallone, L. Ross, and M.R. Lepper. The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, 49, 577-585, 1985.

[33] J.R. Zaller. *The nature and origins of mass opinion.* Cambridge University Press, Cambridge, 1992.

[34] Zaller, *Nature.*

[35] C.S. Taber, M. Lodge, and J. Glather. The Motivated Construction of Political Judgments. In J. Kuklinski, (Ed.), *Citizens and Politics: Perspectives from Political Psychology*, 198-226. Cambridge University Press, Cambridge, 2001.

[36] B. Cronin and S. Pearson. The export of ideas from information science. *Journal of Information Science*, 16, 381-391, 1990.

[37] B. Cronin and L.I. Meho. The shifting balance of intellectual trade in information studies. *Journal of the American Society for Information Science and Technology*, 59, 551-564, 2008.

[38] A.Y.S. Lau and E.W. Coeira. Do people experience cognitive biases while searching for information? *Journal of the American Medical Informatics Association*, 14, 599-608, 2007.

[39] C.A. Mellon. Library anxiety: A grounded theory and its development. *College & Research Libraries*, 47, 160-165, 1986.

[40] H. Poole. *Theories of the Middle Range.* Ablex, Norwood, N.J., 1985.

[41] G. Zipf. *Human behavior and the principle of least effort: An introduction to human ecology.* Addison-Wesley, New York, 1949.

EXPLORING MOTIVES FOR COLLABORATION WITHIN A HUMANITARIAN INTER-ORGANIZATIONAL NETWORK

Louis-Marie Ngamassi
College of Information Sciences & Technology
Penn State University
ltchouakeu@ist.psu.edu

Kang Zhao
College of Information Sciences & Technology
Penn State University
kxz134@psu.edu

Edgar Maldonado
College of Information Sciences & Technology
Penn State University
emaldonado@ist.psu.edu

Carleen Maitland
College of Information Sciences & Technology
Penn State University
cmaitland@ist.psu.edu

Andrea H. Tapia
College of Information Sciences & Technology
Penn State University
atapia@ist.psu.edu

ABSTRACT
While in recent years research has highlighted the rise of inter-organizational collaboration among humanitarian organizations/agencies in the nonprofit sector and has documented issues related to the forming and maintaining of these relationships, there is little known about their motives of collaboration. In this paper, we examine collaboration relationships among organizations/agencies member of a community of interest in humanitarian information exchange. The social network block-model method was used to analyze collaboration network data. Six strongly connected clusters were identified in the community. Evaluating reported reasons for these collaborations, it was found that the two main motivations are relational characteristics, which interestingly are the most and least reported reasons in two of the most densely connected clusters of relationships. These findings suggest that it is important to determine the different reasons for humanitarian - inter-organizational relationships if one is to understand the various patterns of collaboration within inter-organizational networks.

Categories and Subject Descriptors
K.4.m [**Computers and Society**]: Miscellaneous.

General Terms
Human Factors.

Keywords
Inter-organizational network, humanitarian NGOs, social network, collaboration, network clusters.

1. INTRODUCTION
In recent years, nonprofit organizations including those in the humanitarian relief field are increasingly collaborating through alliances, partnerships, and coalitions both within and across sectors [14] [1] [27]. This rise of inter-organizational collaboration is attributed to an increased environmental complexity and challenges where interdependence between different organizations is high and organizational stability is precarious [32] [27]. In the humanitarian sector in particular, as the number of man-made and natural disasters has risen, so has the need for more resources and consequently more collaboration among humanitarian actors [27]. The popularity of these inter-organizational collaborations is well documented in the literature [32] [14] [9]. There is also concurrently, an increasing number of research addressing issues involved in forming and maintaining these inter-organizational collaborations (e.g., [25] [2] [28] [14] [12] [17]).

In their discussion of inter-organization collaboration, Guo & Acar [14] define nonprofit collaboration as what occurs when different nonprofit organizations work together to address problems through joint effort, resources, and decision making and share ownership of the final product or service. The potential gains from inter-organizational collaboration include economic efficiencies, more effective response to shared problems, improvements in the quality of services delivered to clients, the spreading of risks, and increased access to resources [14] [11]. Some studies contend that these network forms enhance organizational effectiveness in ways that traditional governance mechanisms of markets and hierarchies cannot [29] [30]. The advantages offered by network of organizations include greater flexibility and adaptability to change; efficient and reliable information; and reciprocity that can promote long-term stability and reduce uncertainty [29] [13] [33]. Other studies have investigated the potentials drawbacks of inter-organizational collaboration and found that collaboration imposes some costs on partners (e.g. [24] [12]).

In the literature however, little is known about collaboration motives among nonprofit organizations that are members of a

collaboration network. The literature is especially silent on inter-organizational collaboration network of nonprofit in the humanitarian sector. The objective of the paper is to contribute to provide some insights on this aspect of nonprofit inter-organizational collaboration that has been neglected. To this end, we explore collaboration relationships among organizations/agencies member of a community of interest in humanitarian information exchange. Especially, we investigate the patterns of interconnections among organizations/agencies in the community and seek to understand the reasons that explain these collaboration patterns. We conducted a survey among organizations/agencies member of the Global Symposium, a UNOCHA sponsored community of interest on humanitarian information management. The block-model method [20] [6] [35] was used to analyze the data collected. Discussions of the findings draw upon two main concepts including exchange relationship [19] and social network structural equivalence [7] [8] [34] [18].

The rest of the paper is organized as follow: in the following section (Section 2) we present a brief literature review of previous work on inter-organizational collaboration in the nonprofit sector. In section 3 we discuss our analytical framework. Method and data are described in Section 4. The data analysis is presented in Section 5 followed in Section 6 by a discussion and the conclusion.

2. INTER-ORGANIZATIONAL COLLABORATION IN NONPROFIT: LITERATURE REVIEW

As said earlier, researchers have devoted a considerable amount of time investigating inter-organizational collaboration in the specific context of the nonprofit sector (e.g., [14] [9] [24] [1] [12] [16] [17] [11]). They have explored the different forms of collaboration and have looked at the benefits and costs involved in inter-organizational collaboration.

2.1 Forms of inter-organizational collaboration in nonprofit

Studies are also accumulating on the benefits and cost related to inter-organization collaboration in the nonprofit sector (e.g. [9] [16] [24] [12] [17] [11]). Inter-organizational collaboration benefits include benefits to the individual members of the network (e.g. the ability to address shared problems more effectively, the potential for cost savings and organizational learning), benefits to the clients of members of the network (e.g. the higher quality service or end product) and benefits to the community as a whole.

According to Jang & Feiock [17], inter-organizational collaboration among nonprofit organizations has the potential to enhance service to clients. They argue that inter-organizational collaboration is beneficiary to nonprofits because it allows them to share the risks associated with service production and delivery. Gazley [11], identifies five potential gains that nonprofit organizations could ripe from collaborating. They include (i) economic efficiencies, (ii) more effective response to collective problems, (iii) improvements in the quality of services, (iv) the spreading of risks, and (v) increased access to resources. According to Jang [16] collaboration with governments, other nonprofit or private organizations is an attractive option especially when nonprofits face transaction cost.

The major constraints and costs involved in inter-organizational collaboration in the nonprofit sector have also been intensively documented in the literature [12] [27] [23]. They include loss autonomy, financial instability, difficulty in evaluating organizational results, and the opportunity costs from the time and resources devoted to collaborative activities. Nonprofit inter-organizational collaboration must also content with problems related to conflict of interests among organizations and coordination cost in terms of resource inputs, especially staff-time [27]. According to Jang & Feiock [17], the costs of inter-organizational collaboration tend to be individual to organizations that participate in collaborative efforts while the benefits tend to be collective. They assert that nonprofits are confronted with a collective action problem because the benefits of collaborative services are diffused and difficult to measure for individual organizations, but many of the costs are borne by individual organizations.

This vast and growing literature in the nonprofit sector is however silent in investigating the motives of humanitarian inter-organizational collaboration. The objective of this paper is to contribute to the literature by providing some insights on this aspect of collaboration among nonprofit organizations in the humanitarian sector. Our research question is twofold. It is framed as follow: (i) what are the characteristics of interconnections among organizations/agencies which are members of a network of humanitarian information sharing? (ii) What are the major reasons that can explain inter-organizational collaboration patterns observed in a network of humanitarian information sharing?. We discuss below the analytical framework used in the paper. We draw upon network analysis and exchange theory. Network analysis coupled with the theory of exchange provided the framework for our consideration of the relationships within the network. Network analysis captures the embedded nature of a network's organizational actors and structural element [5]. It focuses on patterns of communication and information flows without placing value on the nature of the exchanges. The theory of exchange, meanwhile, assumes that the ties between organizations consist of exchange relations of valued items and that what matters is the value of the items [19] [30]. When combined, network analysis and exchange theory permit to understand more fully the relationships that exist and the nature of these links.

3. ANALYTICAL FRAMEWORK

We use two theoretical lenses to guide our study. These two theories which include the exchange theory and the network structural equivalence are briefly discussed in this section.

3.1 Exchange theory of inter-organizational collaboration

One of the main approaches that inter-organizational researchers have been using to study inter-organizational relationships is the exchange perspective [19] [30]. The exchange theory conceptualizes inter-organizational collaboration more broadly, as to compare with the perspectives of resource dependency and transaction costs theories. This theory posits that organizations get involved in relationships when there is a perception of mutual benefit for interacting. According to Levine & White [19], exchange among organization does not necessarily involve elements of economic value. They assert that part of the exchange process is the development of consensus among organizations. In

addition to explaining the motivations for inter-organizational relationships, the exchange approach also implies that the nature of the interactions between participants in these relationships is characterized by a high level of collaboration [31]. According to Provan & Milward [30], the degree and type of inter-organizational collaboration within a community is reflected in both the number and pattern of inter-organizational exchanges.

3.2 Network structural equivalence

According to the concept of structural equivalence, organizations which have the same or similar ties to others tend to be equivalent in terms of their potential to act in the network [7] [20] [34] [18]. Structural equivalence also takes into account the pattern of connections among all members of the network. Unlike the clique detection methods which are based on relations among members of the sub-group, this approach detects subgroups based on their similar patterns of relations with other members of the network [34] [18]. Members of a network are put in a structurally equivalent group when they have comparable patterns of linkages with other members of the network, even if they do not maintain relations with one another [20].

Central to structural equivalence analysis is the concept of distance [7]. Using the structural equivalence criterion, distance between network members is measured by the degree of similarity in their patterns of interaction: The greater the similarity, the shorter the distance. If two members have exactly identical patterns of relations with other members, their distance from each other is zero. The greater are the differences in their patterns of interaction, the greater is the distance between them. In a nutshell, the goal of structural equivalence analysis is to simplify the structure of relations in a network so that it is possible to understand the various kinds and patterns of interactions occurring in the network.

4. RESEARCH METHODOLOGY

In this paper, we used social network tools to analyze data collected through survey. Network analysis is becoming increasingly popular for understanding complex patterns of relationships. The network perspective examines actors which are connected directly or indirectly by one or many different relationships. Regardless of unit level, network analysis describes structures and patterns of relationships and seeks to understand both their causes and consequences.

4.1 Method

In this paper we analyze data drawn from the Global Symposium inter-organizational project collaboration network [21] [22]. The Global Symposium is a United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) sponsored inter-organizational community for humanitarian information management. The community is made up of about 100 international organizations/agencies, engaged in information management in the field of humanitarian assistance and disaster relief. UNOCHA distinguishes eleven broad categories of network members including NGO, United Nations System, Academia, Donor, Governmental Organization, Regional organization, Intergovernmental Organization, Media, Permanent Mission UN / Observer Private Sector, and Red Cross / Red Crescent Movement. A total of 61 responses were registered from an online survey conducted among 267 attendees of the 2007 Global Symposium+5 meeting. Respondents represented 47 different organizations out of the 119 organizational members of the Global Symposium network that were surveyed; making a response rate of nearly forty percent (39.50%). They were asked to identify organizations/agencies with which they had collaborated on humanitarian projects and to indicate their reasons for collaboration. The survey was the second in a series of three. It was developed with insights gained from survey results obtained at the time of the Symposium itself as well as those gained from an historical analysis of Symposium. Both the first and this second survey were reviewed by leaders of the Symposium. Social network analyses were conducted to explore the data collected in order to assess inter-organizational collaboration patterns in the network. The UCINET software [4] was used to computerize the data. Social network features used in the paper include network density [10] [34], degree centrality [10] [34], network position [7] [8] [34] and a block model [20] [6] [35] [34].

4.2 Data

4.2.1 Project collaboration network data

As said earlier, we collected data through survey, from 47 organizations/agencies members of the Global Symposium. Respondents were asked among other questions, to indicate organizations/agencies with which their organization/agency had collaborated on humanitarian projects. Thirty five (35) organizations answered this question. In order to increase the reliability of this network data, we provided respondents with the complete list of organizations/agencies, rather than relying on their memory. In addition, during coding, we averaged responses from multiple informants of the same project collaboration relationship. Table1 presents the 35*35 directed network matrix generated from the data collected. To protect confidentiality, we identify organizations/agencies by assigning codes for example NGO1. The collaboration relationships represented in the matrix are those reported by organizations on the rows. In this study, we considered both the reciprocated and non-reciprocated reported collaboration ties. A reciprocated collaboration tie is one in which both organizations/agencies report the collaboration relationship. Many researchers report reciprocated ties, with the premise that this strategy increases the reliability of network data and provides a more conservative estimate of inter-organizational relationships (e.g., [26]). However, a relatively high number of non-reciprocated ties are also often reported [3], suggesting that an over reliance on confirmed ties may under represent relationships in the network.

In order to gain a better understanding of tightly and loosely connected members of the network, we used the CONCOR block modeling procedure. CONCOR block modeling method relies on structural equivalence. It aggregates network actors into clusters based on similar patterns of interaction, regardless of whether or not they interact with each other. Table2 shows the matrix resulting from this procedure. The content of this matrix is the same as that of the original network matrix represented by table1. The only difference is that the organizations/agencies in the rows and columns have been reorganized by CONCOR in a manner to group together those that are structurally equivalent. Four different network positions (P1, P2, P3, and P4) are identified. Each position comprises a set of organizations/agencies that collectively reported collaboration or no collaboration with other organizations/agencies in the network.

The CONCOR block modeling procedure also provides a density matrix (Table 3). A density matrix is a table that has positions instead of individual organization/agency as its rows and columns and the values in the matrix are the proportion of ties that are present from the organizations/agencies in the row position to the organizations/agencies in the column position. This density can be used to measure the level of connectedness, which means collaborations in this network, among organizations in the position. In order to define a tightly connected network block, we set the cutoff density value to the density of the whole network which is 0.15. . In other words, a tightly connected cluster is the one in which at least 15% of all possible collaboration ties are effectively made. This method of determining the cutoff density value is frequently used in the literature (e.g. Wasserman & Faust, 1994). Based on this decision, six tightly connected clusters (set of relationships between two positions) were found in the network data. These clusters (P1P2, P2P1, P2P2, P3P1, P3P2 and P4P4) are represented in the image matrix below by 1s (Table 4). The rest of the clusters are represented by 0s.

Table 1. Raw network project collaboration matrix

	NGO1	NGO2	NGO3	NGO4	NGO5	NGO6	NGO7	NGO8	NGO9	NGO10	NGO11	NGO12	NGO13	NGO14	NGO15	NGO16	NGO17	NGO18	NGO19	NGO20	NGO21	NGO22	NGO23	NGO24	NGO25	NGO26	NGO27	NGO28	NGO29	NGO30	NGO31	NGO32	NGO33	NGO34	NGO35
NGO1								1																											
NGO2																																			
NGO3		1				1			1			1				1	1	1	1	1															
NGO4																																			
NGO5			1						1							1	1	1		1															
NGO6			1			1																													
NGO7								1																											
NGO8									1																				1						
NGO9						1				1								1	1																
NGO10		1	1						1	1		1			1			1	1	1	1	1	1	1											
NGO11			1																																
NGO12				1								1																							
NGO13													1	1	1																				
NGO14			1									1																							
NGO15		1	1			1			1			1			1	1	1	1	1	1	1	1	1	1											
NGO16						1										1	1	1	1	1															
NGO17																1																			
NGO18																1																			
NGO19		1			1				1			1	1	1					1																
NGO20												1				1																			
NGO21																																			
NGO22					1				1										1																
NGO23	1											1											1	1											
NGO24				1		1				1				1						1					1										
NGO25		1			1	1									1	1	1		1				1	1											
NGO26				1																															
NGO27		1							1			1	1			1	1	1	1	1	1	1	1	1											
NGO28																													1	1					
NGO29		1	1	1			1	1	1		1	1	1			1	1	1	1	1	1	1	1	1											
NGO30									1																										
NGO31	1																																		
NGO32		1							1			1	1			1	1	1	1	1	1	1	1	1											
NGO33																																			
NGO34									1																										
NGO35																																			

To better understand the collaboration relationship between and within positions, the inter-organizational collaboration network in Table 2 is transferred into the reduced graph in Figure 1. In this graph, positions are represented as nodes and ties between positions in the image matrix define the arcs between nodes. A "1" in an image matrix indicates that there is an arc from the node representing the row position to the node representing the column position in the reduced graph.

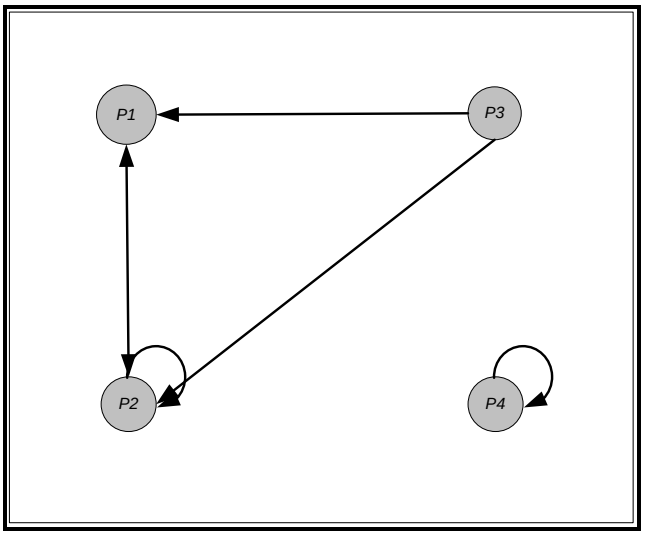


Figure 1. Reduced Graph

Table 2. Blocks of organizations in the network identified through CONCOR block-modeling

		1	9	6	8	2	0	2	2	7	8	4	0	4	1	3	9	5	2	3	2	5	4	5	5	4	0	1	6	3	3	7	1	6	8
1	NGO1								1		1																								
19	NGO19			1		1	1	1				1	1	1																			1		
29	NGO29		1			1	1	1	1	1	1	1	1	1	1	1	1		1		1	1		1	1	1	1	1	1		1				
6	NGO6											1					1																		
8	NGO8													1	1																				
20	NGO20		1																																
22	NGO22		1										1				1																		
32	NGO32		1			1	1		1	1	1	1	1	1	1																			
17	NGO17		1																																
27	NGO27		1	1		1		1		1	1	1	1	1	1	1																		
28	NGO28							1								1																			
4	NGO4		1																																
30	NGO30		1					1					1																						
14	NGO14		1			1						1	1		1																				
31	NGO31	1	1					1		1						1								1	1	1									
33	NGO33		1																																
9	NGO9		1						1				1																						
35	NGO35		1								1					1																			
2	NGO2		1																																
3	NGO3		1			1		1		1			1	1	1	1	1																		
12	NGO12		1	1		1		1								1																			
5	NGO5		1										1	1	1	1	1																		
24	NGO24		1														1				1											1		1	1
15	NGO15		1			1		1			1	1	1		1	1					1			1	1	1									
25	NGO25		1						1	1	1			1			1					1				1					1				
34	NGO34		1											1									1			1									
10	NGO10		1			1		1	1	1		1	1		1	1					1		1	1	1										
11	NGO11		1											1							1														
16	NGO16		1				1				1		1	1	1		1																		
13	NGO13		1						1			1		1																					
23	NGO23	1	1	1									1		1								1		1		1								
7	NGO7																																1		
21	NGO21																																		
26	NGO26																				1												1		
18	NGO18		1																																

4.2.2 Data on reasons for collaboration

Respondents to the survey were also asked to indicate the reasons their organizations/agencies collaborate with other organizations/agencies member of the network. They were provided with a list of eight reasons, derived from the literature on coordination in general but tailored to the specific concerns of humanitarian information management (Table 5), from which they could select all that apply to them.

Table 3. Density Matrix

	P1	*P2*	*P3*	*P4*
P1	0.100	**0.400**	0.147	0.050
P2	**0.218**	**0.264**	0.024	0.000
P3	**0.240**	**0.297**	0.110	0.033
P4	0.050	0.000	0.017	**0.167**

Table 4. Image Matrix

	P1	*P2*	*P3*	*P4*
P1	0	1	0	0
P2	1	1	0	0
P3	1	1	0	0
P4	0	0	0	1

Table 5. List of reasons for collaboration

R1	The goals of both organizations overlap.
R2	The project was on my organization's agenda already.
R3	Both organizations are operating in the same geographical area.
R4	My organization is seeking a relationship with the project partner.
R5	The other organization has a successful track record of securing project funding.
R6	The other organization has data in which my organization is interested.
R7	The other organization has information management policies or procedures in which that my organization is interested.
R8	The other organization has technical tools in which that my organization is interested.

Table 6 shows the responses that were collected. These responses were aggregated for each of the six tightly connected network clusters identified through CONCOR. The aggregation was made based on the number of reported project collaboration relationships in each cluster. For example, if organization/agency NGO1 collaborates for reason R1, this reason will be credited with the total number of collaborations report by NGO1. As said earlier, we assumed that all reported collaborations from one organization/agency were for the same reasons. After calculating the total frequency of occurrence of each reason, we computed the mean frequency per cluster (Table 7) and ranked them from the most important (high mean frequency) to the least important (low mean frequency). Table 8 presents the result of the ranking.

Table 6. Organizations' reasons for collaboration

		R1	R2	R3	R4	R5	R6	R7	R8
P1	NGO1	√							
	NGO19	√					√	√	√
	NGO29	√	√	√	√	√	√	√	√
	NGO6	√	√		√	√			
	NGO8	√		√			√		√
P2	NGO20			√			√	√	√
	NGO22	√	√	√	√		√	√	
	NGO32	√	√	√	√		√	√	√
	NGO17	√							
	NGO27	√		√	√		√	√	√
	NGO28	√							
	NGO4	√	√			√			√
	NGO30	√	√	√	√	√	√	√	√
	NGO14	√	√	√	√	√	√	√	√
	NGO31	√					√		√
	NGO33	√	√	√				√	√
P3	NGO9	√				√			
	NGO35	√	√	√			√	√	√
	NGO2				√	√			√
	NGO3	√						√	
	NGO12	√	√	√	√	√	√	√	√
	NGO5				√		√		√
	NGO24	√	√		√	√	√	√	√
	NGO15	√	√	√	√	√	√		
	NGO25	√	√		√	√	√	√	√
	NGO34	√	√			√			
	NGO10	√			√	√			
	NGO11	√					√	√	√
	NGO16	√	√		√	√			√
	NGO13	√						√	√
	NGO23	√			√	√	√		√
P4	NGO7	√					√	√	√
	NGO21	√	√		√		√		
	NGO26				√	√	√		
	NGO18		√			√	√		√

5. ANALYSIS

5.1 Characterizing network positions

As shown in table 3, applying the CONCOR procedure to the network data produced four structurally equivalent positions in the network. The number of organizations/agencies in each the network positions varies significantly ranging from 4 (four) to 15 (fifteen). Positions P1 and P4 have the smallest number of organizations/agencies, 5 (five) and 4 (four) respectively. These two positions could also be characterized as NGOs positions since 4 (four) out of the 5 (five) organizations/agencies in position P1 and 2 (two) out of the 4 (four) in position P4 are NGOs. Position P2 in made up of 11 (eleven) organizations/agencies mainly from the UN System (six out of eleven). The only Donor organization in the 35 surveyed belongs to this position. This position could be characterized as the UN position. Position P3 has the greatest number of organizations/agencies (fifteen) and is the most diversified in term of different categories represented (eight). With six organizations/agencies, academia is the category with the highest number of organizations/agencies. The only Media organization surveyed belongs to this position. Position P3 could be characterized as the 'other agencies' position. This examination of the Global Symposium collaboration network positions sheds some light on the grouping of the members of the network.

Table 7. Mean frequency each network position reported types of reasons for collaborations

Cluster	R1	R2	R3	R4	R5	R6	R7	R8
P1-P2	1.07	1.03	1.37	0.78	0.93	1.19	1.32	1.23
P2-P1	0.98	0.94	1.55	0.72	0.43	1.03	1.14	1.18
P2-P2	1.07	1.04	1.92	1.19	0.35	1.28	1.41	1.27
P3-P1	0.95	0.84	0.52	0.95	1.23	0.84	0.76	1.03
P3-P2	0.96	0.85	0.55	0.98	1.10	0.70	0.74	0.72

Table 8. Ranking of types of reasons for collaboration in descending order of mean frequency

	Cluster				
	P1-P2	*P2-P1*	*P2-P2*	*P3-P1*	*P3-P2*
1	R3	R3	R3	R5	R5
2	R7	R8	R7	R8	R4
3	R8	R7	R6	R4	R1
4	R6	R6	R8	R1	R2
5	R1	R1	R4	R2	R7
6	R2	R2	R1	R6	R8
7	R5	R4	R2	R7	R6
8	R4	R5	R5	R3	R3

5.2 Patterns of collaboration

After the network is partitioned into structurally equivalent positions, patterns of relationships between and within the positions are examined using the density matrix and the image matrix (see [51(p.389-391)]). As said earlier, a density matrix shows the proportion of potential linkages that are actually sent from a row position to a column position. It is possible for a position to send many linkages to other positions and not to receive linkages in return. Another possibility is for a position to be internally linked, with members of the block sending links to one another.

Six tightly connected clusters of collaboration were identified in the Global Symposium network data. With regards to the density of interactions, these clusters present diversified patterns of project collaboration between and within the four structurally equivalent network positions. Scores in the density matrix range from 0.40 to 0.167. For example, the cluster formed by positions P1P2 is strongly linked. Forty percent (40%) of all the possible linkages between the organizations in these positions are actually found to exit. In contrast, only about 17% of all possible linkages of organizations/agencies in the cluster formed by positions P4P4 are present.

5.2.1 Patterns of collaboration within positions

Among the six tightly connected clusters of interaction that were identified in the network data, two were concerned with interaction within position (P2P2 and P4P4). The level of collaboration among organizations/agencies in each of these two positions was higher than the average in the whole network. These two clusters differ in their intensity of interaction as well as in the type of organizations/agencies. With 26.4% of connections, P2P2 has one of the highest densities among the tightly connected clusters while P4P4 has the lowest density. Position P2 is made up mainly of UN agencies while P4 is composed of NGOs. The reduced graph (Figure 1) shows that P4 is an isolate in the network. That is, organizations/agencies in this position collaborate only among themselves.

5.2.2 Patterns of collaboration between positions

The following four clusters of interaction P1P2, P2P1, P3P1, and P3P2 show collaboration between network positions. An examination of the direction of relationship flows between positions in the reduced graph (Figure 1) shows a "one way" relationships between positions P3 and P1 and positions P3 and P2. This means that organizations/agencies in position P3 reported collaboration with organizations/agencies in both position P1 and position P2. But organizations/agencies in P1 and P2 did not report collaboration relationships with P3. This may be a common characterization of relationships between resources providers and resources seekers. The pattern of relationships is consistent with this notion. The reduced graph also shows a bidirectional relationship between position P1 and position P2.

5.3 Reasons for collaboration

Table 7 shows for each of the six tightly connected clusters of interactions the mean frequency of occurrence of reasons for collaboration. The highest score cross cluster is for reason R3 (both organizations are operating in the same geographical area). This highest score is registered in cluster P2P2. The lowest score cross cluster is for reason R5 (the other organization has a successful track record of securing project funding). This score is also registered in the cluster P2P2. The appearance of these two extremes scores in the same cluster would be a strong indication

of the characteristic of interaction in this cluster. This also indicates the differences between clusters.

An examination of table 7 also shows that two different main reasons for collaboration (highest scores) are identified that could characterize two of the six clusters. As said earlier, reason R3 would characterize cluster P2P2, while R5 (Successful track record of securing project funding R5) would characterize cluster P3P2. These two reasons occupy respectively and inversely the top and the last positions in the two clusters. This same pattern is almost similar in the other clusters.

Table 9 presents the density of collaboration among organizations grouped per reported reasons of collaboration. As highlighted in this table, R3, R4, R5 and R7 register respectively the first, second and third highest density of collaboration. These findings are concordant with the result obtained from block modeling.

6. DISCUSSION AND CONCLUSIONS

The main objective of this research is to investigate inter-organizational collaboration behavior / reasons among humanitarian organizations/agencies which are members of a community of interest in information exchange. We seek to understand the patterns of interconnections among organizations/agencies in the community. We also investigate the reasons that explain the collaboration patterns observed in the community. Although previous research highlight the popularity of inter-organizational collaboration in the nonprofit sector and document issues involved in forming and maintaining these inter-organizational collaborations few studies examine the behavior of humanitarian organizations/agencies members of a community of interest in information sharing.

Table 9. Density of collaboration among organizations grouped per reason

	Reasons for Collaboration							
	R1	R2	R3	R4	R5	R6	R7	R8
# of organizations	30	17	12	18	17	22	17	23
Max # of ties possible	870	272	132	306	272	462	272	506
# of ties present	166	93	77	124	103	128	103	124
Density	0.19	0.34	0.58	0.41	0.38	0.28	0.38	0.25

Our study shows that with regards to inter-collaboration relationships, the UNOCHA Global Symposium community is fragmented into four groups described as network positions. The density of collaboration relationships within and between these groups varies significantly ranging from 0% (zero) to 40% (forty). Organizations/agencies of each group appear to be almost all in similar category (e.g. NGO, UN agencies, Academia). This may mean that organizations in similar categories hold similar structural positions in the inter-organizational collaboration network. The study also shows that two main reasons predominantly characterize collaboration relationships among members of the Global Symposium community. These reasons were related to (i) location of operation, i.e., both organizations/agencies are operating in the same geographical area and (ii) resources i.e., the other organization has a successful track record of securing project funding. More importantly, we found that the two predominant reasons were inversely the most and least reported in two of the most densely connected clusters. These findings are consistent with Bolland & Wilson [3] according to whom every inter-organizational network is clustered into groups of agencies centered on specific needs. Our study extends their work in the humanitarian information exchange field.

As stated earlier, according to the exchange perspective of inter-organizational relationships, relations form when organizations perceive mutual benefits or gains from interacting [19] [15]. Our findings corroborate with this perspective as one of the major reasons for which organizations collaborate was found to be related to secured resources. When looking at the findings from the structural equivalence perspective [7] [8] [34], the fact that the two predominant reasons for collaboration were inversely the most and the least reported in two different clusters would be consistent with this approach. Organizations in the same structurally equivalent network position would tend to have similar behavior in the network. The results of this research contribute to the body of literature inter-organizational collaboration among humanitarian organizations/agencies by identifying and describing the patterns of collaboration as well as the motives that could explain these patterns.

Summarizing, this paper responds to a call for researchers to further examine solutions to inter-organizational collaboration issues. It sheds some lights on collaboration behavior in a community of interest in humanitarian information exchange. It also identifies some factors that explain the patterns of collaboration found in the community.

The results of this study should be considered in light of several limitations. Of particular concern, is the potential sampling bias due to the fact the survey participants were not selected through any scientific sampling technique. Rather, the survey was conducted on a sample defined by UNOCHA thereby generating an organizational bias. Another limitation to the study concerns the source of information. The network data was constructed based on information provided by individuals. The position of these individuals in their organization may not allow them to always have complete information about the organization's relationships. A third limitation concerns CONCOR, the social network block model that we use. CONCOR has been criticized as lacking validation. That is, there is no proof that convergence of the correlation matrix actually represents structurally equivalent positions. Lastly, two important assumptions are made in the study. First, we assume that inter-organizational collaboration relationships are of different kind. At any particular time, an organization could be engaged collaboratively in different kind of projects with other organizations. The second assumption is that reasons for which an organization collaborates with others were the same irrespective of projects or collaboration partner's characteristics.

7. ACKNOWLEDGMENTS

This work was partly supported by National Science Foundation grant number CMMI-0624219.

8. REFERENCES

[1] Arya, B., & Lin, Z., 2007. Understanding Collaboration Outcomes From an Extended Resource-Based View Perspective: The Roles of Organizational Characteristics, Partner Attributes, and Network structures. Journal of Management, 33(5), 697-723

[2] Austin, J. E. 2000. Strategic collaboration between nonprofits and business. Nonprofit and Voluntary Sector Quarterly, 29(1). 69-97.

[3] Bolland, J. M., & Wilson, J. V. 1994. Three faces of integrative coordination: A model of interorganizational relations in community-based health and human services. Human Services Research, 29(3), 341–365.

[4] Borgatti S. P., Everett M. G., Freeman L. C. 1999. UCINET 6.0 Version 1.00. Natick: Analytic Technologies.

[5] Brass, D. J., Galaskiewicz, J., Greve, H. R., & Tsai,W. 2004. Taking stock of networks and organizations: A multilevel perspective. Academy of Management Journal, 47, 795-817.

[6] Breiger, R.L., Boorman, S.A. and Arabie, P. 1975. An algorithm for clustering relational data, with application to social network analysis and comparison with multidimensional scaling. Journal of Mathematical Psychology, 12, 328-383.

[7] Burt, R.S. 1976. Positions in Networks. Social Forces, 55(1), 93-122.

[8] Burt, R.S. 2008. Information and Structural Holes: Comment on Reagans and Zuckerman. Industrial and Corporate Change, 17(5), 953-969.

[9] Feiock, R. C. and Andrew, S. 2006. Introduction: Understanding the Relationships Between Nonprofit Organizations and Local Governments. International Journal of Public Administration, 29(10&11), 759-767.

[10] Freeman, L. C., 1979. Centrality in Social Networks: Conceptual Clarification. Social Networks, 1, 215-239.

[11] Gazley, B., 2008. Beyond the contract: The scope and nature of informal government nonprofit partnerships. Public Administration Review, 68(1), 141-154

[12] Gazley, B., and Brudney, L., 2007. The purpose (and perils) of government-nonprofit partnership. Nonprofit and Voluntary Sector Quarterly, 36(3), 389-415

[13] Gulati, R. 1995. Social structure and alliance formation patterns: A longitudinal analysis. Administrative Science Quarterly, 40, 619-652.

[14] Guo, C., & Acar, M. 2005. Understanding Collaboration Among Nonprofit Organizations: Combining Resource Dependency, Institutional, and Network Perspectives. NonProfit and Voluntary Sector Quarterly 34(3), 340-361.

[15] Hall, R., Clark, J., Giordano, P., Johnson, P., & Van Roekel, M. 1977. Patterns of Interorganizational Relationships. Administrative Science Quarterly, 22(3), 457-474.

[16] Jang, H. 2006. Contracting Out Parks and Recreation Services: Correcting for Selection Bias Using Heckman Selection Model. International Journal of Public Administration. 29 (10&11). 799-818.

[17] Jang, H and Feiock, R., 2007. Public and Private Funding Reliance of Nonprofit Organizations: Implications for Inter-organizational Collaboration. Forthcoming in Public Productivity and Management Review.

[18] Kilduff, M., & Tsai, W. 2003. Social Networks and Organizations. London: Sage.

[19] Levine, S., & White, P., 1961. Exchange as a conceptual framework for the study on inter-organizational relationships. Administrative Science Quarterly, 5: 583-601.

[20] Lorrain, F., and White, H.C. 1971. Structural equivalence of individuals in social networks. Journal of Mathematical Sociology 1: 49-80.

[21] Maitland, C. and Tapia, A. 2007. "Outcomes from the UN OCHA 2002 Symposium & HIN Workshops on Best Practices in Humanitarian Information Management and Exchange." Prepared for United Nations Office for the Coordination of Humanitarian Affairs, October 13th, 2007. 16 pages.

[22] Maitland, C. and Tapia, A. 2007b. "Global Symposium +5 Information for Humanitarian Action Survey Report." Prepared for United Nations Office for the Coordination of Humanitarian Affairs, November 15, 2007. 33 pages.

[23] Maitland, C., Ngamassi, L. & Tapia, A. 2009. Information Management and Technology Issues Addressed by Humanitarian Relief Coordination Bodies . Proceedings of the 6th International ISCRAM Conference – Göteborg, Sweden, May 2009

[24] McGuire, M. 2006. Collaborative public management: Assessing what we know and how we know it. Public Administration Review. 66(1). 33-43.

[25] Milne, G. R., Iyer, E. S., & Gooding-Williams, S. 1996. Environmental organization alliance relationships within and across nonprofit, business, and government sectors. Journal of Public Policy & Marketing, 15(2), 203-215.

[26] Morrissey, J. P., Calloway, M., Bartko, W. T., Ridgley, S., Goldman, H. H., & Paulson, R. I. 1994. Local mental health authorities and service system change: Evidence from the RobertWood Johnson program on Chronic Mental Illness. Milbank Quarterly, 72(1), 49–80.

[27] Ngamassi, L., Maldonado, E., Zhao, K., Robinson, H., Maitland, C., and Tapia, A. 2008. Exploring Barriers to Coordination between Humanitarian NGOs: A Comparative Case Study of Two NGO's Information Technology Coordination Bodies. Under review at the International Journal of Information Systems and Social Change (IJISSC), special issue on IS/IT in Nonprofits.

[28] O'Regan, K. M., & Oster, S. M. 2000. Nonprofit and for-profit partnerships: Rationale and challenges of cross-sector contracting. Nonprofit and Voluntary Sector Quarterly, 29(1), 120-140.

[29] Powell, W. W. 1990. Neither market nor hierarchy: Network forms of organization. Research in Organizational Behavior, 12, 295-336.

[30] Provan, K. G.,&Milward, H. B. 1995. A preliminary theory of inter-organizational network effectiveness: A comparative

study of four community mental health systems. Administrative Science Quarterly, 40(1), 1-33.

[31] Schmidt, S., & Kochan, T., 1977. Interorganizational relationships: Patterns and Motivations. Administrative Science Quarterly, 22(2), 220-234

[32] Stone, M. 2000. Exploring the Effects of Collaborations on Member Organizations: Washington County's Welfare-to-Work Partnership. Nonprofit and Voluntary Sector Quarterly, 29, 98-119.

[33] Tang, J., 2007. Effect of Information Networking on Organizational Coordination: Case Study of an Electronic Travel Community. International Journal of Management, 24(4): 774-789.

[34] Wasserman, S., & Faust, K. 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.

[35] White, C., Boorman, A., & Breiger, R. 1976. Social Structure from Multiple Networks. Blockmodels of Roles and Positions. The American Journal of Sociology, 81(4), 730-780.

Information Horizons of Taiwanese Graduate Students

Tien-I Tsai
Doctoral student, School of Library and Information Studies, University of Wisconsin-Madison
Rm 4160A, Helen C. White Hall, 600 N. Park St., Madison, WI 53706, USA
ttsai5@wisc.edu

ABSTRACT

The information horizon is an imaginary field that users position their information sources according to their perceived importance. Previous research investigated Internet users' information horizons and pointed out that the information source accessibility and quality play an important role in the information horizons and orient people's information seeking behavior. This study examined how the perceived source accessibility and perceived source quality influence Taiwanese graduate students positioning their information sources in their information horizons. The study aims to examine the influence of perceived source accessibility and quality on students' information horizons, and to examine the differences of information horizons among students from different disciplines. Two methods were employed to collect data: the information horizon map drawn by Taiwanese graduate students and interviews with those students. Nine Taiwanese graduate students at University of Wisconsin-Madison were recruited. Results showed that all students tend to include more information sources in the center (most preferable) than the peripheral zone (least preferable) on their information horizon map. However, students from humanities and social sciences included more information sources in their information horizons than students from sciences did. Contrary to previous information horizon research on everyday information seeking behavior, this study showed that despite the fact that graduate students from all disciplines preferred information sources with high accessibility, they also considered quality as an important factor. Future research may focus on a specific concept of information horizons, such as social networks, and include different groups of International students and compare with their American counterparts to learn more about students' information horizons under research contexts among disciplines and cultures.

Keywords

Information Horizon, Information Behavior, Source Accessibility, Source Quality, Taiwan

1. INTRODUCTION

Articles about information seeking behavior constitute a relatively large portion of Library and Information Science research. With the rapid development of the Internet, researchers have been investigating the influence of the Internet on users' information seeking behavior as well as the principle of least effort in their information behavior. In recent years, more attention has been paid to contextual, situational, or role variables (e.g. students, patients, etc.) than usual demographic variables (e.g. age, gender, etc.), and these studies have a common concern with sources and channels – typically interpersonal channels [2]. However, while humans' information behavior may be easy to learn, it is indeed a rather intricate and perplexing issue. Information behavior can be influenced by a variety of factors. Not only demographic variables but also contextual, situational, and role variables are critical for analyzing and understanding humans' information behavior. So studies which consider both demographic and role variables may be even more valuable.

In a perspectivist viewpoint, spatial factors in information seeking are perceived by users and thus are highly subjective and depend on contexts and situations [15]. Sonnenwald developed a theoretical framework of information horizons to depict the concepts of contexts, situations, and social networks in humans' information behavior [19]. She defined information horizon as how information seekers perceive the usefulness of information sources. In an information horizon map, users positioned information sources according to their perceived importance in various contexts. In this vein, Savolainen and Kari investigated Internet users' information horizons in a non-work role situation and pointed out that information seeking is oriented by the information source horizons [14]. They found that the accessibility and quality of information sources/channels play a particularly important role. Therefore, information horizons may not only show the contexts, situations, and networks in users' information behavior, but also exhibit users' perceived information accessibility and quality. Thus studies on information horizons may expand the knowledge of users' information need, information seeking and information use as a whole.

Sonnenwald claimed that information horizons research have influenced studies on information needs of high school students, graduate and undergraduate students, older adults, and professionals [20]. Nevertheless, very few studies have been actually based on the theoretical framework of information horizons, and what studies have been conducted are mainly about information seeking behavior of non-work purposes [8, 14, 16]. Although recent studies by Huvila on information horizons moved the focus on work roles, no article investigated graduate students' information behavior in terms of their information horizons. In addition, articles investigating graduate students information behavior in terms of spatial factors did not focus on their information horizons [4, 5]. Regarding the lack of knowledge in graduate students' information horizons, it would contribute to the field if we explore more about graduate students' information horizons and integrate their information needs, information seeking and information use – in terms of information source selection – into this framework. Given the researcher's background as a Taiwanese international student, she believes that conducting an exploratory study on a group of international students that she is most familiar with would be a good start.

Therefore, this study examined how the perceived source accessibility and perceived source quality may influence Taiwanese graduate students positioning their information sources in their information horizons. The purpose of this study included: 1. to examine the influence of perceived source accessibility on students' information horizons; 2. to examine the influence of perceived source quality on students' information horizons; and 3. to examine the differences of information horizons among students from different disciplines. Three research questions were

raised: First, what do Taiwanese graduate students include in their information horizons? Second, how do they determine what to include in their information horizons in terms of perceived source accessibility and quality? And third, how do the information horizons differ among Taiwanese graduate students' in different disciplines? The ultimate purpose of this research was to get a better understanding about the information horizons of Taiwanese graduate students in research context so that libraries may provide more suitable services to assist graduate students in doing their research. Specifically, exploring students' most preferred information sources and channels may provide some basic guidance for collection development or collection management decisions. Exploring what information sources or channels were least preferred by students may provide implications on library services promotion or information literacy programs.

2. LITERAURE REVIEW

Information behavior studies related to spatial factors can be categorized into three approaches according to the levels of abstraction: the objectifying approach, the realistic-pragmatic approach, and the perspectivist approach [15]. Examining users' information horizons is exploring information behavior with a perspectivist approach. Briefly introducing these different approaches may help us picture where information horizons are situated in information behavior research.

In the article "Spatial Factors as Contextual Qualifiers of Information Seeking, "Savolainen described and discussed the three approaches [15]. Traditionally, researchers use the objectifying approach to discuss spatial factors in information behavior in a physical sense, and view spatial factors as something discrete and entity-like which may constrain information seeking. In a realistic-pragmatic approach, information seekers may rethink the role of spatial factors and redefine their source preferences by abandoning time consuming visits to a remote library and search for information on the Internet instead. This approach is derived mainly from the revised information pathways by Johnson [7], Pettigrew's information grounds [12], and Chatman [3]'s small world theory. Finally, the perspectivist approach also shares the same proposition with the realistic-pragmatic approach that spatial factors not only constrain information seeking, but also enable it. However, the perspectivist approach focuses more on the subjective and situation-bound interpretation of spatial factors. The theory of information horizon mainly constitutes the perspective approach in information seeking research. This perspective approach provides a viewpoint for examining how people subjectively assess the significance of information sources and spatially construct their information horizons.

Diane Sonnenwald proposed the concept of information horizon and defined it as a map user positions information sources according to their perceived importance in various contexts [15]. This concept of information horizon provides the analysis of source preferences with a conceptual framework [14]. Despite proposing a theoretical framework, Diane Sonnenwald also provides a basic guideline for the research design of information horizons [20]. She points out that the purpose for studying information horizons is to examine when and why people access/ not access individuals or other information resources, to examine relationships among information resources, and to examine the proactive nature of information seeking process. The methods usually include semi-structured interviews with a critical incident interview technique and a map-drawing technique.

Empirical studies based on the theoretical framework of information horizons are mainly about the daily life of people and the non-work purposes. Several articles follow this semi-structured interview method, and some adapt the critical incident technique to investigate people's information seeking behavior. For example, Savolainen and Kari conducted a research study on information seeking by Internet users in the context of self-development [14]. They asked users to place the information sources in three zones on the map of information horizons according to their preferences. The most preferred sources were placed in Zone 1, while the least preferred sources were placed in Zone 3. In their study, the perceived source quality was given less attention in Zone 1, and the perceived accessibility, as well as quality, were weighed equally in Zone 2. In Zone 3, the information selection criteria mainly focused on the perceived source quality, and the assessments were easily affected by the negative experiences. They revised the definition of information horizon proposed by Sonnenwald, and defined it as "an imaginary field, which opens before the 'mind's eyes' of the onlooker – information seeker" [14]. Broadly speaking, these horizons can be defined as a perceived information environment. However, Huvila shifts the discussion on information horizons to work roles. Based on this conceptual framework, he discussed the work and work roles through a task-oriented approach [4]. Following his perspective, this current study looks into information horizons by viewing students and their coursework-related projects as work roles and tasks. In another study, instead of employing user-drawn information horizon maps, Huvila introduces an "analytical information horizon maps (AHIM)" and further develops a new method to approach users' information horizons [5]. The information horizon maps are drawn by the researcher according to the data collected from the interviews.

Other related studies on students' information behavior in terms of spatial factors might not be based on the theoretical framework of information horizons. There is indeed very little known about the information horizons of undergraduate or graduate students. For instance, Lee examined the students' interactions with information environment and how the structure of the information environment affected students' information seeking behavior [10]. Her study provided linkage among information seeking, information organization, and collection development. However, although Lee utilized a similar concept of presenting her interview findings on a map with immediate space, adjacent space, and outside space to show an information space for college students, her study is conducted in a more pragmatic approach with physical spatial factors [10]. Based on this pragmatic approach, this current study viewed the information space in a perspective view and tried to explore the information horizons of graduate students. Based on Lee's information space [10], Tsai investigated the citing behavior of Library and Information Science graduate students in National Taiwan University and found a similar information space which include online resources and personal collection in the immediate space, interpersonal channels and nearby libraries in the adjacent space, and interlibrary loan services, other libraries and bookstores in the outside space [21].

In addition, with the increasing international student population in the United States, there are articles exploring topics on academic library services and multicultural communities; however, fewer articles focus on the international students and study their

information needs and information seeking behavior [11], let alone the information horizons of a certain group of international students. Previous research shows that some library services may be new to international students and they may encounter linguistic, cultural, and technological barriers with their library use, which may also depend on their previous library experiences [11, 18]. Therefore, this current study examines the information horizons of Taiwanese graduate students studying in the United States in order to delve into and understand the information behavior of this specific group of international students.

3. THEORETICAL FRAMEWORK

The theoretical framework of information horizons proposed by Sonnenwald in 1999 was based on empirical studies of information behavior from various settings, specifically, Kuhlthau [9]'s information seeking model, Wilson [22]'s general model of human information behavior, as well as Belkin [1]'s and Ingwersen [6]'s studies [19]. Rather than providing specific factors for predicting certain changes in information seeking behavior, Sonnenwald believed that understanding information behavior as a process is more important, and examining the role of three fundamental concepts – context, situations, and social networks – in the theoretical framework of information horizons helps us explore information behavior as a process. In the framework of information horizons, contexts are multi-dimensional and can be described by attributes including place, time, goals, tasks, systems, situations, processes, organizations, and types of participants. Additionally, a flow of situations constitute a context. A situation may be characterized as a set of related activities, or a set of related stories that occur over time; however, situations within any given context are not necessarily linearly-ordered discrete events. Social networks refer to communication among individuals, in particular, patterns of connection and resonance interaction. Social networks help construct and are constructed by situations and contexts. This theoretical framework contains five propositions that describe the relationships of the three concepts stated above [19, p.181-188]:

- Human information behavior is shaped by and shapes individuals, social networks, situations, and contexts.
- Individuals or systems within a particular situation and context may perceive, reflect, and/or evaluate change in others, self, and/or their environment.
- Within a context and situation is an "information horizon" in which we can act.
- Human information seeking behavior may, ideally, be viewed as collaboration among an individual and information resources.
- Because information horizons consist of a variety of information resources, many of which have some knowledge of each other, information horizons may be conceptualized as densely populated spaces.

The information horizons provide a theoretical basis by proposing three concepts and five propositions for an evolving framework of information exploration, seeking, filtering, use, and dissemination. Additionally, this framework of Information horizons was built on previous research in information studies, communication and psychology [19]. Although this framework did not indicate how to design effective strategies for enhancing information seeking, it conceptualized the three fundamental concepts to describe information behavior, incorporates cognitive, social, and system perspective, and aims at providing implications for system design [19].

However, while emphasizing the importance of social networks, Sonnenwald did not include information resources in the definition of the three fundamental concepts – contexts, situations, or social networks. It is reasonable to emphasize social networks in information horizons, but the information resources seem to play one of the critical roles in the information horizons, so that she mentions information resources in two of the five propositions. Instead of limiting the third concept to social networks which refer to individual members within situations and contexts, the third concept might be modified as network so that we can incorporate the information resources in this concept. Therefore, this study focused on three concepts – contexts, situations, and networks (rather than the narrowly defined "social network"). Moreover, based on previous research, Savolainen and Kari investigated Internet users' information horizons and pointed out that the information source accessibility and quality play an important role in the information source horizons and orient people's information seeking behavior [14]. This study tried to extract different elements that constitute the perceived information source accessibility and quality and explored how the two factors affect students' positioning of information sources on their horizon maps.

As for the methodology, Diane Sonnenwald provided a basic guideline for the research design of information horizons [20]. She pointed out that to learn how users position their information resources the methods usually include semi-structured interviews with a critical incident interview technique and a map-drawing technique. The map of information horizon which shows all information resources, including people, provides graphical articulation of the information horizon in a particular context. And the interview provides verbal articulation of the information horizon in that context. These methods not only help to describe the information resources used, but also explain their importance and role in the information seeking process.

However, most empirical studies based on this framework do not include a map drawing technique. One of the reasons might be it is difficult for subjects to include all the channels they used in a certain context or situation. Furthermore, as Sonnenwald mentioned, we should understand information behavior as a process and view information selection within those complex contexts as a dynamic process. Therefore, it might be difficult to approach such complex issues with a single map. Map drawing technique could give us a clearer picture of what information horizons are about. Nevertheless, there are still issues whether or not we can approach such complicated information behavior through a static map. An alternative way to employ a map drawing technique was a new methodology by Huvila. Huvila discussed the drawbacks of the original map drawing by Sonnenwald and proposed an analytical information horizon maps (AIHM) [5]. He believes that it would be more beneficial to draw a map from the interview records by the researcher rather than ask the participants to draw a map without interviewing them. Through this way, the researcher can structure and analyze typical information behaviors better. However, information horizon maps derived from this method would be totally interpreted by the researcher and thus might not actually explore the initial inquiry of learning how users position the information sources on their maps.

In order to gain a better understanding from the users' perspective, this study employed both interviews and map-drawings to examine the information horizons of Taiwanese graduate students in research contexts. Specifically, the focus of this study was on the research-related task situations in which students use or not use certain resources, and on students' information source networks in terms of their information needs and information seeking behavior.

4. METHODS

4.1 Data Collection

The study employed two methods for data collection: the information horizon map drawn by Taiwanese graduate students and semi-structured interviews with those students. The former may provide a clear picture about the information horizons of the students, while the latter may provide rich data for examples and explanations about how they position information sources on the maps. Additionally, this triangulation may increase the validity of this research.

There are approximately 150 Taiwanese graduate students at UW-Madison. The study recruited Taiwanese graduate students at UW-Madison through the mailing list of the UW-Madison Student Association of Taiwan (UWSAT). A consent form addressing the purpose and design as well as confidentiality was given to the participant when recruiting. For the exploratory nature of qualitative study, smaller sample size is required to obtain an in-depth understanding of certain phenomena, and thus this study recruited nine volunteers to participate. A purposive sampling was employed to balance the demographics of the sample, i.e., graduate students from humanities, social sciences, and hard sciences were equally sampled.

Nine Taiwanese graduate students volunteered to participate in this study. Four are master's students and five are doctoral students. Their age ranged from 25 to 33. Among which, three students are from each discipline, i.e. humanities, social sciences and hard sciences. Their majors include: English, music, linguistics, educational psychology, counseling psychology, consumer science, electrical and computer engineering, mechanical engineering, and civil and environmental engineering. The number of years they have been to the United States ranged from 1 to 5 (See Table 1).

Table 1. The demographic characteristics of the participants

Participant	Gender	Master's/ Doctoral	Discipline	Number of Years Studying in the U. S.	Task
Hannah	F	D	Humanities	4	Dissertation
Hank	M	D	Humanities	2	Thesis
Helen	F	M	Humanities	2	Final project
Sadie	F	M	Social Sciences	2	Final project
Selena	F	M	Social Sciences	1	Final project
Sandra	F	M	Social Sciences	1	Final project
Charles	M	D	Hard Sciences	5	Dissertation
Chris	M	D	Hard Sciences	1	Research project
Craig	M	D	Hard Sciences	5	Lab project

Table 1 also shows the tasks in research contexts that participant recalled in the interviews. Despite two students who have been studying in their current programs for 4 or 5 years recalled their dissertation, other students mainly recalled their final projects or lab projects as the research contexts. These specific tasks provided a basis of students' information horizons and determined their information needs. According to Sonnenwald, the doctoral dissertation task could be a situation and the department, discipline, state of art, etc. could be contexts [19]. Therefore, these tasks graduate students mentioned can be viewed as the situations while their discipline and level of degree can be viewed as the contexts in the information horizons. The networks of the information horizons in the findings would be how the information sources and channels related to one another. The nodes of the networks would be those information sources and channels, including the document types graduate students used, the people they consulted, and the places or ways they accessed the information sources.

In the semi-structured interview, a critical incident technique was employed to help graduate students recall their information needs and information source selection experiences. Participants were asked to recall the process of their research projects or theses/ dissertations in terms of what information sources they use and through what channels they access these resources, especially what are their information source preferences for their research related tasks. Participants were then asked to draw an information horizon map showing what information sources they use and through what channels they access these resources for course related tasks (e.g. doing a final project or paper, writing thesis or dissertation, etc.). They positioned these information sources in three given zones according to their preference (See Appendices 11.1). Afterwards, they were asked to explain their rationales, how important each resource or channel is to them when doing those tasks, and provide examples of situations they would access/use each resource. Additionally, they were asked to describe if there is anyone or any information source that lead them to other people or information sources. Each individual interview (including map drawing) ranged from 30 minutes to an hour.

Although Taiwanese graduate students studying at UW-Madison read English, it would be much easier for them to read their native language. Therefore, the researcher translated the instruction for map drawing and the interview guide from English to traditional

Chinese. In order to maintain the validity of this study, the researcher asked another Mandarin-speaking graduate student with high proficiency in English to look at both versions of the instrument and make sure the Chinese version were similarly translated. Please see appendices 11.1 for the instrument (English version).

4.2 Data Analysis

The interview was audio recorded and then transcribed for analysis. Pseudonyms were assigned to every participant in order to maintain confidentiality. The researcher analyzed the transcripts by coding related concepts in the transcripts into perceived accessibility and perceived quality to provide possible explanations of the students' information horizons. NVivo 8 was used as an analysis tool. Data collected from the interview transcripts are analyzed in descriptive, topic, and analytical levels. According to Richards [13], descriptive coding which informs the attributes of cases, e.g. person's gender, may also occur in quantitative studies. Topic coding merely allocates passages to topics which involves little interpretation. Analytical coding requires interpretation from descriptive and topic coding. The researcher conducted descriptive coding with the casebook in NVivo to provide the background information of participants and also conducted some topic coding to provide information sources/ channels mentioned by participants. Furthermore, the researcher also conducted analytical coding with tree nodes and matrix queries in NVivo, based on the research questions to provide possible explanations for the information source positioning of Taiwanese graduate students. For example, the statements related to advisors, colleagues, or friends were separately labeled as child nodes and were placed under the parent node of interpersonal channels. A code book is developed through the process of data analysis stated above. Specifically, following the five propositions in the information horizons, the researcher analyzed the background of individual as well as the situations and networks mentioned by the participant to see how they shape individuals' information behavior. The researcher analyzed how individual perceive, reflect, and/or evaluate change in others, self, and/or their environment by examining their rationale about access or not access/ use or not use certain information sources.

Data collected from the maps was analyzed by descriptive statistics adopted by Savolainen and Kari and was then further analyzed by comparing the results among different disciplines to see the similarities and differences [14]. In Savolainen and Kari's study, they calculated the number of sources/channels in each zone. They also weighted sources and channels by multiplying a source/ channel by 3 in Zone1, 2 in Zone 2, 1 in Zone 3, to see the weighted scores of each sources or channels. The researcher followed this analytical method and examined the distribution of types of information sources in the given three zones, and then analyzed the similarities and differences among disciplines.

5. RESULTS

5.1 Information Sources on the Information Horizon Maps

According to the results of this study, nine participants included 133 information sources or channels on their information horizon maps. Among which, 54 sources are in Zone 1, 46 sources are in Zone 2, and 33 sources are in Zone 3 (Table 2). In order not to inflate the number of different sources or distort the result of the source distribution in the three zones, sources or channels that mentioned several times were only counted once, and thus derived 114 different sources were derived from the maps. Overall, humanities students listed the most sources while hard science students listed the least (See Appendices 11.2). All students listed the least number of sources in Zone 3. Moreover, the patterns of the distribution of the sources in the three zones for social science and hard science students were similar. The more central the zone is, the larger the number of sources is.

Table 2. Distribution of sources/channels in different zones and disciplines

Number of Sources/ Channels	Humanities	Social Sciences	Hard Sciences	Total
Zone 1	17	21	15	53
Zone 2	22	16	8	46
Zone 3	16	13	4	33
Total	55	50	28	133

The researcher listed all information sources/ channels mentioned by graduate students in zone 1 to zone 3 (Table 3 to 5 shows the sources mentioned by graduate students in zone 1 to 3 accordingly), and categorized information sources/ channels into Internet, personal collections, interpersonal channels, library collections, media, lab resources, bookstores, other libraries/public libraries, and bibliography. Specific information sources under each category were listed as what participants wrote on the maps.

Table 3. Research information sources in zone 1[1]

Discipline	Humanities	Social Sciences	Hard Sciences
Source Type	Internet E-resource (MLA), Library online catalog, [Document delivery services], Department library website links, Grove online dictionary, Amazon Personal collections Personal collection Syllabus Interpersonal channels	Internet Databases (3), Library online catalog (2), Google (2), Open Access Journals, Reports from research institutes, Google Scholar (conference, working paper, journal articles), Government publications, [Document delivery services] Interpersonal channels	Internet Conference proceedings (2), Google, Google Scholar (2), Databases (3) (IEEE, ACM, Web of Science), Document delivery services Interpersonal channels Advisor (2), Lab cooperative partner (from other compan

[1] The numbers in the parentheses are the number of times participants positioned that information source

Discipline	Humanities	Social Sciences	Hard Sciences
	Advisor, Professors, Friends Library collections Journals (Articles), Theses, Books, Audio CD, Library reference collections, Scores from school	Professors (2) Personal collections Library collections E-journals, E-books, Books, Articles Media TV, News	y) Lab resources Theses, Instruments Library collections Journals

Table 4. Research information sources in zone 2[2]

Discipline	Humanities	Social Sciences	Hard Sciences
Source Type	Internet Google (review, thesis, score analysis), Databases (2), Wiki (not really reliable), Academic society website (need permission), Personal collections Scores, Pamphlets from Audio CD Interpersonal channels Taiwanese Classmates (research area), Classmates, Librarians, Master class feature artists, Studio class peer review, Audio materials from classmates Library collections	Media TV programs Interpersonal channels Lab colleagues (suggestions, information), Friends, Professor, Classroom peers (2) Library collections Reference books (2) (Dictionaries and others), Books (2) Internet Wiki, Amazon, Online Catalog, Document delivery services, Online News Personal Collections	Internet Google Interpersonal channels Classmates Library collections Journals (2), Textbooks, Books (2) Lab resources Software

[2] The numbers in the parentheses are the number of times participants positioned that information source

Discipline	Humanities	Social Sciences	Hard Sciences
	Printed journal, Books, Interlibrary loan (ILL) Media History channel (discovery), Movies, TV series, Radio programs Bookstore (Bookman, Eslite bookstores)		

Table 5. Research information sources in zone 3[3]

Discipline	Humanities	Social Sciences	Hard Sciences
Source Type	Internet Google Library collections Journals, Magazines (2), Other articles, Bibliography in books Bookstores Bookstore, Buy scores Other libraries, Public libraries Interpersonal channels Friends (2), Classmates, E-mail professors for resources, Writing center instructors	Library collections Print journals, Books (2), Theses, Magazines, Microfilms Interpersonal channels Friends (2), Family, Writing center instructors Media Radio programs, Newspapers, TV News	Library collections Theses, Books Interpersonal channels Advisor, Colleagues

Weighting each source/ channel by multiplying the number of informants mentioned by 3 in zone 1, 2 in zone 2, 1 in zone 3 helps us see students' preferences. Table 6 to Table 8 shows the weighted scores in each zone. Tables 9 to 11 show the weighted scores of research information resources mentioned by students from each discipline. All of the graduate students preferred sources through the Internet and tend to position most information sources in zone 1 and zone 2. However, graduate students from different disciplines have different preference for positioning

[3] The numbers in the parentheses are the number of times participants positioned that information source

different sources in each zone. Compared to humanities and social science graduate students, hard science students tend to use fewer sources in each zone mainly through the Internet, interpersonal channels, lab resources, and library collections. On the other hand, humanities and social science students tend to use a variety of sources through the Internet, interpersonal channels, personal collections, library, and so on. Nevertheless, humanities students uniquely mentioned purchasing books or other research materials while social science students mentioned getting research ideas through the media.

Table 6. Weighted scores of research information sources in zone 1

Source Type	Humanities	Social Sciences	Hard Sciences	Total
Internet	18	39	27	84
Library collections	18	12	3	33
Interpersonal channels	9	6	9	24
Personal collections	6	3	N/A	9
Media	N/A	6	N/A	6
Lab resources	N/A	N/A	3	3

Table 7. Weighted scores of research information sources in zone 2

Source Type	Humanities	Social Sciences	Hard Sciences	Total
Interpersonal channels	12	10	2	24
Library collections	6	8	10	24
Internet	10	10	2	22
Media	8	2	N/A	8
Personal collections	4	2	N/A	6
Bookstores	2	N/A	N/A	2
Bibliography	2	N/A	N/A	2
Lab resources	N/A	N/A	2	2

Table 8. Weighted scores of research information sources in zone 3

Source Type	Humanities	Social Sciences	Hard Sciences	Total
Library collections	4	6	2	12
Interpersonal channels	5	4	2	11
Media	N/A	3	N/A	3
Bookstores	3	N/A	N/A	3
Other libraries/Public libraries	2	N/A	N/A	2
Internet	1	N/A	N/A	1
Bibliography	1	N/A	N/A	1

Table 9. Weighted scores of research information sources mentioned by humanities students

Source Type	Zone 1	Zone 2	Zone 3	Total
Internet	18	10	1	29
Library collections	18	6	4	28
Interpersonal channels	9	12	5	26
Personal collections	6	4	N/A	10
Media	N/A	8	N/A	8
Bookstores	N/A	2	3	5
Bibliography	N/A	2	1	3
Other libraries/Public libraries	N/A	N/A	2	2
Total	51	44	16	111

Table 10. Weighted scores of research information sources mentioned by social science students

Source Type	Zone 1	Zone 2	Zone 3	Total
Internet	39	10	N/A	49
Library collections	12	8	6	26
Interpersonal channels	6	10	4	20
Personal collections	3	2	N/A	5
Media	6	2	3	11
Total	66	32	13	111

Table 11. Weighted scores of research information sources mentioned by hard science students

Source Type	Zone 1	Zone 2	Zone 3	Total
Internet	27	2	N/A	29
Library collections	3	10	2	15

Source Type	Zone 1	Zone 2	Zone 3	Total
Interpersonal channels	9	2	2	13
Lab resources	3	2	N/A	5
Total	42	16	4	62

5.2 Perceived Accessibility and Quality as Factors in Positioning Information Sources on the Maps

In this study, the perceived accessibility refers to the ease students perceive in accessing an information source. Based on interviews, properties of the perceived accessibility include convenience, efforts needed, time needed, familiarity, flexibility, etc. Convenience and efforts needed are possibly the most obvious elements in accessibility. Many studies have raised the issue of the principle of least effort in users' information behavior [2]. This study also demonstrates parallel findings. Taiwanese graduate students also place high emphasis on convenience and tend to try whichever they consider the convenient way to access an information source, especially under the time pressure.

Sadie (Soc, M)[4]: Convenience is the most important thing because toward the end of the semester, all you want is get your finals done, so the convenient information sources make me happy.

Hank (Hum, D): The library service is very convenient. You can request the book to your preferred library, which is very nice. Although it makes people lazy, it's really nice.

Selena (Soc, M): I think the library online system is very convenient, so I use it a lot.

Students also regard efforts and time needed as important factors. Students prefer the sources that are perceived not far from them and do not require much effort. They think the faster the sources can be accessed, the less effort they need to make. And it seems that students' information horizon maps start from the computer and themselves, reaching other people and resources.

Selena (Soc, M): [To access sources in] zone 2 need some efforts. For example, you need to walk to someone or something to access [the information source]. For this [sources in zone 1], you only need to sit down and you can start on your own.

Hannah (Hum, D): I usually try the Internet before the library because it is faster. It is easier.

Additionally, lengthy books may also intimidate students. Under time pressure, students tend to read an article or a chapter, instead of a whole book. Some students admit their difficulties or laziness and, to some extent, are intimidated by books. Although reading English books would be much slower than Chinese books, they also clarify that no matter the books are in English or Chinese, they feel almost the same way.

[4] The parenthesis after the participant's name reminds readers his/her discipline and education level by using abbreviations: Hum for humanities, Soc for social sciences, Sci for hard sciences; M for master's level, and D for doctoral level.

Sadie (Soc, M): Because I'm lazy, I accept to read one chapter, but a whole book... I know I don't have time to read, so I put it in zone 3…If it [the book] was in Chinese, I might be a little more willing to read. However, I won't read the whole book, either.

Selena (Soc, M): When I pick up a book, it's hard for me to read from the first page. Because it [the book] is too thick, you don't know where to start…

Other important elements in accessibility may encompass familiarity and flexibility. Students usually turn to information sources that they are familiar with. Unfamiliar information sources may be in the outer field of their information horizons. Sources they do not even know may be excluded unconsciously.

Hank (Hum, D): I am familiar with other ways [to access the information source], so I access the resources in my familiar ways. I don't do something that I am not familiar with and stumble around.

Selena (Soc, M): I did not use or I didn't want to use certain resources probably because I don't know them. Probably once I know how to use them, I will love them.

Craig (Sci, D): I used journal articles a lot because it is easier to find journal articles. You can use the Web of Science. It's hard to find a conference paper.

Students also mention that they prefer a source they can access on their own, rather than relying on others, since it is more flexible to work on their own. This reiterates the previous point that students perceived sources that can be accessed by using a computer requires less effort, and thus their horizon map starts from the Internet and online resources.

Sadie (Soc, M): [When searching on the web], you can look up in a dictionary, you don't feel you're wasting people's time, and you don't worry about how long it will take or how fast you can read. No one monitors you. I think that's more flexible and comfortable.

Perceived quality refers to the characteristics of information sources that students consider relevant or suitable to use. Based on the interviews, the properties of perceived quality include authority and credibility, helpfulness, publication date (if it is recently published), and relevance. Authority and credibility are important attributes that graduate students emphasize. When the students mention the online resources to which they refer in their papers, they usually assess the authority and credibility of the Website by its reputation or credentials of the author and/or the institution of that Website. Graduate students mentioned that they take reliability and academic style, for example, into consideration when prioritizing information sources or deciding whether or not the source would be appropriate to be included in their papers. Additionally, not only can one assess authority and credibility based on the creator or institution of that website, authority and credibility can also be assessed by interactions with individuals.

Helen (Hum, M): [I used it because] it was a website of a professor. If it is a website of nobody, I won't use it. I will see the credibility [and decide whether I'll include the resource or not].

Sadie (Soc, M): A professor has authority. He knows what academics want. He has more experiences, so he knows what you need to include [in your paper]… These are what the Websites cannot offer you, so I think they are supplementary.

Selena (Soc, M): I believe if I can find it [from the library], it is reliable… I would like to make it look academic, and the databases from the library are helpful. For example, the resource

I need is a fairy tale, so I go to the library [online catalog], I can find some [original] children's books, rather than other versions that have been adapted in a movie or something.

Some graduate students mentioned that they sometimes determine the preference or importance of information sources by its helpfulness to their research. When it comes to information sources that they are not sure are helpful, they tend not to regard the source as top priorities. Similarly, when asking people questions about their research, students tend to ask the ones they think have enough expertise to be helpful to their research.

Sadie (Soc, M): My dad sometimes tells me how to write my thesis. He'd say: you need to have a goal and a question, and blah, blah, blah…He shows me his master's thesis. I think it could be a resource, but I don't know if it would be helpful…So parents' opinions may be another resource, I don't know if it is helpful though.

Charles (Sci, D): Professors are definitely important to us, but colleagues are only sometimes helpful…Everyone has his/her own research interests.

Sadie (Soc, M): I think the wording of your questionnaire is important and they [colleagues] are Americans, so I ask them to look at my questionnaire and change the wording for me.

Additionally, recently published materials are especially preferred by science students. Although all of the science students showed concerns about the accuracy of conference papers, they all put very high premium on recently published journal articles. On the contrary, all science students mentioned the outdated books are not very useful for their research.

Charles (Sci, D: Actually, we don't need to survey a lot of papers because our research topics are usually about very new ideas. For example, the area of my research topic starts from 2001, so I cannot go further before 2001 [I can probably find nothing on my literature before 2001]…The most highly used [material] is conference proceeding because it is the latest one. The next would be the journal or transaction, and the last would be the textbook because it takes so long to publish a book. It might take several years.

Chris (Sci, D): Books are usually too old, so they are not valuable for writing papers for publications. We would prefer journals or conference papers. Conference papers are newer, but they sometimes make mistakes. Journals are usually peer-reviewed.

Craig (Sci, D): I think journal papers are more accurate [than conference papers]. Sometimes conference papers are only about people's research process. You may get some new ideas from them, but they are not done yet.

As for social science students, they preferred recently published materials under some circumstances but did not stress as much as science students. Social science students mentioned that for some topics related to media or policies it is very important to include the up-to-date information so as to get an in-depth understanding about the topic. However, they also admitted the importance of some classic books or articles, especially on certain crucial concepts or theories.

Sadie (Soc, M): I prefer recently published journal articles, especially on the topic of adolescence's media use, because the media change so fast. I think the more recent ones [articles] are more helpful.

Chris (Sci, D): I think theses and dissertations are more helpful [than books] because it is his own research, the author knows it very well.

Finally, relevance is also an important property of perceived quality. Students emphasize that the information they need is the relevant materials to their research topic. Students may do the known item search for a specific author or title on the search engine, the library Website, or the database, especially when they know the important scholars, articles, or books in their research area. When they think the source they found is highly relevant to the topic, they often use strategies like snowball techniques to gather more relevant information for their papers.

Helen (Hum, M): I'll put it [personal collections] in Zone 2 because I'm not sure if it is directly related to my topic, and I don't know if it is reliable… Personal owned materials tend to be more general, but writing a paper should be more specific. So I'm afraid it wouldn't be closely related [to my paper because what I owned are basically textbooks]. I'll see if I have directly relevant materials [personal collections]. If I think I owned something really relevant, I'd definitely go find it.

6. DISCUSSION

6.1 Information Horizons and Source Positioning Considerations for Students from Different Disciplines

All graduate students prefer sources through the Internet, including search engines, databases, library online catalogs, and so on. However, graduate students from different disciplines have different preferences for information sources and channels. Compared to humanities and social science graduate students, hard science students tend to use fewer sources, mainly through the Internet, interpersonal channels, lab resources, and library collections. On the other hand, humanities and social science students tend to use a variety of sources through the Internet, interpersonal channels, personal collections, library, and so on. Nevertheless, humanities students uniquely mentioned purchasing books or other research materials while social science students mentioned getting research ideas through the media.

As for the information selection, graduate students of all disciplines not only prefer sources with high accessibility, but also are concerned with the quality of the sources. Although the Internet is always highly preferred by students in all disciplines, humanities and social science students mentioned their quality concerns. They also tend to judge the source quality when they use those highly accessible resources in order to maintain a certain level of credibility. However, contrary to Savolainen and Kari's finding on Internet users' information seeking, interpersonal channels in research contexts was not perceived as an easily accessible source or channel in general [14]. Students mentioned that familiarity and friendliness may influence their perceived accessibility to that interpersonal channel. In addition, students also mentioned that even if professors are nice, they think professors are extremely busy, which sometimes impedes them from asking professors questions. All these perceived factors may constitute graduate students' positioning of the information sources in their information horizons.

As for perceived accessibility per se, students prefer whatever is online regardless of where the actual source comes from. For instance, even if personal collection would be physically close to

them, some students prefer online because it does not limit the source to a certain place like their "home" or their "office." Another example would be the use of document delivery services. Almost all students mentioned the convenience of the Library Express service.[5] Some students point out that they do not care whether or not the article is in a locally owned journal simply because it is very convenient to use Library Express to place a request. That is, students care more about the convenience of the information channel through which they access an information source, rather than where the information source came.

Interestingly, hard science students tend to judge an article by the times cited as well as the academic reputation of the author while humanities and social science students don't. Although the Internet is always placed in Zone 1 by students in all disciplines, it is also placed in Zone 3 by humanities students due to quality concerns. Contrary to the findings from Savolainen and Kari on everyday information seeking, the current study did not show a clear difference in students' selection criteria among the three zones [14]. For example, there may not be a clear difference between the selection criteria in Zone 1 and Zone 2. One of the possible reasons might be the different context. For everyday information seeking, people tend to place whatever is accessible in Zone 1. However, students tend to use what is convenient to them when doing research. In the meanwhile, they also tend to judge the source quality when they use those highly accessible resources in order to maintain certain level of credibility. Therefore, there are no obvious differences in information selection by perceived accessibility and quality in Zone 1 and Zone 2.

Overall, students across the disciplines agreed that their positioning criteria for Zone 1 are mainly based on convenience and perceived usefulness or importance. Some students mentioned they also consider the familiarity and flexibility. A tendency of including both perceived accessibility and perceived quality are salient. Information sources positioned in Zone 2 were those that needs some efforts, provides more general rather than specific ideas or definitions (sometimes not directly cited in the paper), are not that frequently used, or resources that may not be very reliable. Finally, information sources that were positioned in Zone 3 were those with no urgent needs, may trouble others, are too expensive, and require the most efforts (need quality control, are too far, need to read a lot to get only few ideas, need to be prepared before asking others).

6.2 Networks in the Information Horizons for Taiwanese Graduate Students

Networks as one of the fundamental concepts in students' information horizons could also be discovered from the results in this study. The networks in students' information horizons in this study can be generally divided into two categories: social networks which started from interpersonal channels and resource networks which started from information sources other than interpersonal channels.

One of the important network structures is the social networks starting from interpersonal channels. Interpersonal channels such as professors, senior colleagues and colleagues usually refer students to other useful information sources, including prestigious scholars, articles, books, or online resources. Almost all students talked about the suggestions from the professor. For example, Selena mentioned that sometimes the professor may suggest you to read some journal articles or something that he has read before. However, an noticeable outlier in sciences mentioned that he thinks his advisor is too busy and has no time to talk to him, so he sometimes emails the professor only to report his progress, not asking questions. And thus, unlike other Taiwanese students, he placed the advisor in zone 3. This may be another aspect of the perceived accessibility concern.

An interesting finding here is that Taiwanese graduate students tend to specify senior colleagues from their peers in the same class year, senior colleagues may be considered more experienced and thus be one of the good interpersonal information sources. For example, Sadie mentioned that sometimes senior colleagues would suggest you to look at someone's articles, or suggest you add some articles on certain topic to make [your literature review] more complete.

Another important network structure is the resource networks which usually start from bibliographies of important sources. Most students mentioned bibliographies from a highly relevant article or book usually lead them to a vast amount of other useful information sources. Among the students, Hank pointed out that "there are a lot of bibliographies in the Norton [textbook]. Those bibliographies were edited by professors, so I found it as treasures that can help you find [useful materials] very quickly." Other interesting networks could start from previous syllabi, relevant theses, or online resource such as Google Scholar. For instance, Charles mentioned that "before someone published his/her thesis, he/she usually published something in other format. So I'd try to use the author to search other articles."

6.3 The Effects of Language Issues and Previous Experiences in the Information Source Positioning for Taiwanese Graduate Students

Some other interesting findings of the study come from the concerns of international students. Although writing center instructors were not positioned in the center of students' information horizons (positioned in Zone 3), humanities and social science students tend to mention writing center instructors while hard science students don't. Humanities and social science students mentioned that consulting with writing center instructors either helps convey their ideas more clearly in English or helps them with academic English usages.

Other language issues include being afraid to cite Chinese works in English, being afraid to repeatedly ask people questions due to language barriers, and so on. For example, Selena mentioned that she is afraid of translating what she has read in Chinese and citing it in her paper since she may not translate or summarize the passage accurately, and she does not know the citation format of citing works in other languages. However, she also mentioned the difficulties when she knows she read something in Chinese but does not know how to find an article or book in English that talks about the same idea. Sadie mentioned worries about asking people questions in person. She said that "you know he [the professor] is busy and you don't want to bother him too much, so sometimes when you don't really understand his answers, you don't want to ask him again. If you browse on the Web, you can look up in a

[5] Library Express is the UW-Madison's document delivery and interlibrary loan service.

dictionary, and you don't feel guilty as you are not wasting other people's time."

Students also have concerns about the price and fees of services, based on their previous experiences in Taiwan. Hannah pointed out that "since whatever you buy here [in the United States] are much more expensive than in Taiwan, so I hardly ever purchase books or scores here." Selena admitted that "I don't know if there is a fee for the [document delivery] service, so I didn't use that service. And I am not sure if that article is a highly related one."

In sum, Taiwanese graduate students are concerned with some language issues and naturally relate their previous experiences in Taiwan. These concerns can also be viewed as another aspect of perceived accessibility which may affect their information source positioning and thus influence their information horizon maps.

7. LIMITATIONS

Several limitations of this study exist. First, the result of this study will not be able to be generalized to a larger population. Due to the small population of Taiwanese graduate students at UW-Madison and the small sample size for the qualitative research, there were only nine participants in this study. In addition, although the study sampled graduate students according to their disciplines, students from sciences recruited in this research were all from engineering related fields. No pure science students participated in this research. Furthermore, participants were all from the same institution, and their education level did not match the discipline well. All social science students were in master's programs, while all hard science students were in doctoral programs. Moreover, the retrospective design relied on students recalling their research related activities. However, students may not recall all the information sources in their coursework related activities, which may have influenced the results of the present study. Finally, although this study had a second coder for the instrument translation, were it be a second coder for data analysis, it would help increase the validity of this study.

8. CONCLUSION

Information behavior studies have proliferated with the rapid development of the Internet, shifting the focus from traditional demographic variables to contextual, situational and role variables. This study tried to incorporate the two sets of variables, exploring the similarities and differences among students from different disciplines within situations under research contexts. Since people's information horizons orient their information seeking behavior, discovering users' information horizons help understand users' information seeking behavior. We may expand our knowledge of what information sources and channels graduate students prefer or perceive as important to them, and why they tend to use or tend not to use certain sources or channels. Thus, this study not only contributes a better understanding of graduate students' information horizons when seeking information in a research context, but also contributes to the scarce literature on graduate students' information horizons in the field of information behavior, especially in academic librarianship. Additionally, discovering the similarities and differences among disciplines may not only serve as a whole picture of the information horizons for graduate students, but also help illustrate the different nature of disciplines. Although all students tend to include more information sources in the center (most preferable) than the peripheral zone (least preferable) on their information horizon map, students from humanities and social sciences included more information sources in their information horizons than students from sciences did since science students tend to do a newer topic and focus on his/her own experiment. Hard science students also tend to judge an article by the times cited and/or the academic reputation of the author due to their needs for replicating experiments and their demands for obtaining accurate results.

Furthermore, according to Liao, Finn, and Lu, international graduate students from a unique multicultural user group for the university libraries. Understanding and meeting their needs will help them achieve higher level of academic success and enhance universities' teaching and research capacities [11]. Viewing Taiwanese graduate students as a group of international students with multicultural backgrounds may help find out the characteristics of international students' information needs in terms of positioning information sources and channels in their horizon maps under research contexts. This study could be a starting point for investigating different groups of international students' information horizons.

All of the above may yield some implications for libraries to provide more suitable services to assist graduate students in doing their research. For example, libraries can promote services that students may not know, e.g. interlibrary loan services, document delivery services or request a purchase. In addition, libraries can also provide library orientations or workshops on different databases. In sum, this study may fill the gap in knowledge of information behavior research and shed light on library services as well. Moreover, from this research graduate students may learn their information behavior as well as possible information sources that they can use in their research.

Future research may focus on a specific concept of information horizons, such as social networks, and include different groups of International students and compare with their American counterparts. More interesting findings may also be elicited if we shift our focus on the interpersonal channel since it is an important component of social networks in students' information horizons and there might be underlying cultural differences which influence the interpersonal channels in students' information horizon maps under research contexts. Incorporating social network theory and social network analysis would probably help us gain a more in-depth understanding of students' information horizons under research contexts, and learn more about the similarities and differences among disciplines and cultures.

9. ACKNOWLEDGMENTS

I would like to express my thanks to Associate Professors Kyung-Sun Kim, Ethelene Whitmire, Kristin Eschenfelder, Assistant Professor Catherine Arnott Smith, and three reviewers for providing me with comments and suggestions on this paper.

10. REFERENCES

[1] Belkin, N. J. Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, 5(05 1980), 133-143.

[2] Case, D. O. 2007. Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. Emerald, Bingley, UK.

[3] Chatman, E.A. 1991. Life in a small world: applicability of gratification theory to information-seeking behavior. Journal of the American Society for Information Science, 42, 6, 438-449.

[4] Huvila, I. 2008. Work and work roles: a context of tasks. Journal of Documentation, 64, 6 (12 2008), 797-815.

[5] Huvila, I. 2009. Analytical information horizon maps. Library & Information Science Research, 31, 1 (01 2009), 18-28. DOI=10.1016/j.lisr.2008.06.005.

[6] Ingwersen, P. Cognitive perspectives of information retrieval interaction elements of a cognitive IR theory. Journal of Documentation, 52, 1 (03 1996), 3-50.

[7] Johnson, J. D. 1996. Information seeking: an organizational dilemma. Westport, CN: Quorum Books.

[8] Kari, J. and Savolainen, R. 2003. Towards a contextual model of information seeking on the Web. New Review of Information Behaviour Research, 4, 1 (12 2003), 155-175. DOI=10.1080/14716310310001631507.

[9] Kuhlthau, C. C. 2004. Seeking Meaning: A Process Approach to Library and Information Services (2nd ed.). Westport, CT: Library Unlimited.

[10] Lee, H. 2003. Information spaces and collections: Implications for organization. Library & Information Science Research, 25, 419-436.

[11] Liao, Y, Finn, M, and Lu, J. 2007. Information-seeking behavior of international graduate students vs. American graduate students: A user study at Virginia tech 2005. College & Research Libraries, 68, 5-25.

[12] Pettigrew, K.E. 1999. Waiting for chiropody: contextual results from an ethnographic study of the information behaviour among attendees at community clinics. Information Processing & Management, 35, 6, 801-817.

[13] Richards, L. 2005. Handling Qualitative Data: a Practical Guide. Thousand Oaks, CA: Sage.

[14] Savolainen, R, Kari, J. 2004. Placing the internet in information source horizons. A study of information seeking by internet users in the context of self-development. Library & Information Science Research, 26, 415-433.

[15] Savolainen, R. 2006. Spatial factors as contextual qualifiers of information seeking. Information Research, 11, n.p.

[16] Savolainen, R. 2007. Information source horizons and source preferences of environmental activists: A social phenomenological approach. Journal of the American Society for Information Science & Technology, 58, 1709-1719.

[17] Savolainen, R. 2008. Source preferences in the context of seeking problem-specific information. Information Processing & Management, 44, 274-293.

[18] Song, Y. 2005. A comparative study on information-seeking behaviors of domestic and international business students. Research Strategies, 20, 23-34.

[19] Sonnenwald, D. H. 1999. Evolving perspectives of human information behaviour: Contexts, situations, social networks and information horizons, 176-190.

[20] Sonnenwald, D. H. 2005. Information Horizons. In Fisher K. E., Erdelez, S., and McKechnie, L. E. F. Theories of Information Behavior. Medford, NJ: Information Today, Inc.

[21] Tsai, T-I. 2008. The Influence of Information Availability on Citations of These and Dissertations in Library and Information Science. Unpublished Master's Thesis. National Taiwan University. [in Chinese]

[22] Wilson, T. D. The cognitive approach to information-seeking behavior and information use. Social Science Information Studies, 4, 2 (04 1984), 197-204.

11. APPENDICES

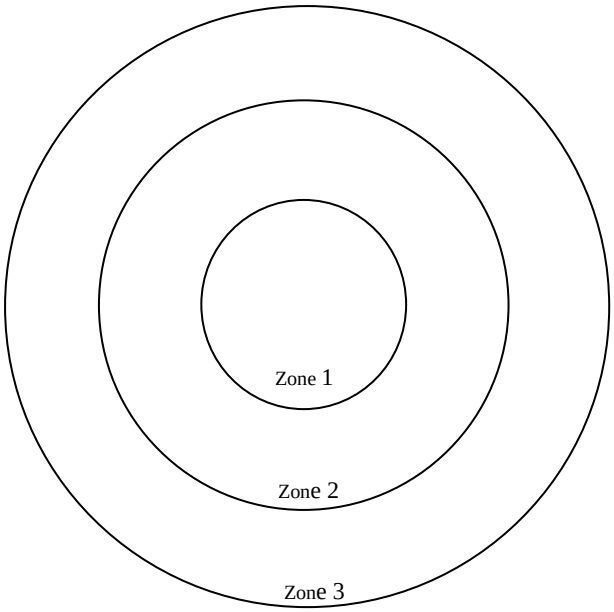
11.1 Instruments

11.1.1 Interview Guide

- Please recall your recent research paper/ project or any other research that is important/ meaningful to you. E.g. your term paper, your thesis or dissertation, etc. And describe your information selection and use experience from the beginning to the end. Can you divide your research process into different stages in terms of information selection and use?
- What do you usually do when gathering resources at the very beginning of your research? (E.g. searching online databases, talking to people, etc.)
- What information resources do you access in each stage of your research? What resources do you prefer in each stage? In what situations do you usually access or not access those resources?
- Among these resources, what do online resources mean to you in each stage of your research?
- Where have you been to access information resources during your research process? Where do you prefer to access information resources in each stage of your research? In what situations do you usually go to certain places to access those resources?
- Among these resources and people, are there any resources or people lead you to other resources or people you access/use?
- Who do you consult with in each stage of your research? Who do you prefer to consult with when you have questions or problems regarding your research? In what situations did you usually consult with or not consult with people?
- Please describe any experience about any resource or people that you once thought about but didn't have a chance to access or use. And explain why you did not access or use it.

11.1.2 Instructions for drawing your information horizons

Please draw a map describing your information source preference for course related tasks (e.g. doing a final project or paper, writing thesis or dissertation, etc.). Please try to include all the resources you use for course related tasks and write the resources you prefer the most in Zone 1, the second preferred resources in Zone 2, and the least preferred resource in Zone 3.

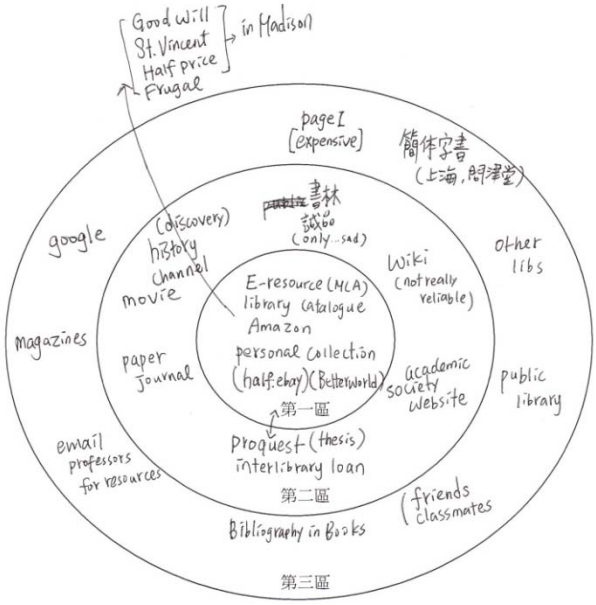


Please explain the resources and channels you include in your information horizon map by giving examples from your course/research related experiences (e.g. doing a final project or paper, writing thesis or dissertation, etc.).

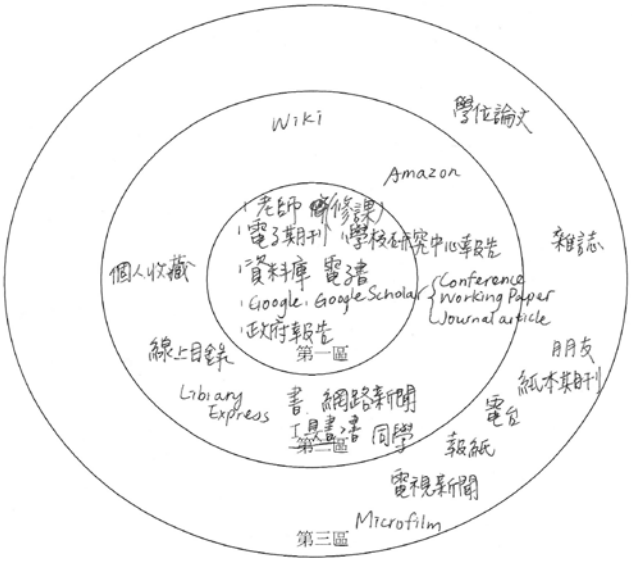
- Among the resources included in your map, what resources do you prefer the most? And why?
- Among the resources included in your map, what resources do you prefer the least? And why?
- Among the people included in your map, who did you consult with when doing your course related tasks? Who did you prefer to consult with when you have questions or problems regarding to these tasks? And why?
- How do you prioritize among the sources and channels shown on your map? What do you usually take into consideration when you access or not access those resources or channels?
- Is there any resource or channel that you once thought about but didn't have a chance to access or use it (which is not shown on your map)? Please describe the situation.

11.2 Information Horizon Maps (Examples)

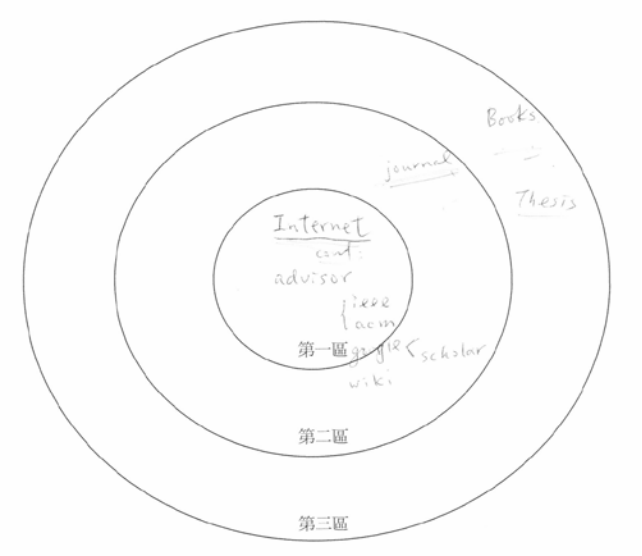
Map from Humanities Student



Map from Social Science Student



Map from Hard Science Student


Books
journal
Thesis
Internet
advisor
ieee
acm
第一區
scholar
wiki
第二區
第三區

Building an IT Taxonomy with Co-occurrence Analysis, Hierarchical Clustering, and Multidimensional Scaling

Chia-jung Tsui, Ping Wang, Kenneth R. Fleischmann, Asad B. Sayeed, and Amy Weinberg
University of Maryland, College Park, MD 20742
{ctsui, pwang, kfleisch, asayeed, weinber}@umd.edu

ABSTRACT
Different information technologies (ITs) are related in complex ways. How can the relationships among a large number of ITs be described and analyzed in a representative, dynamic, and scalable way? In this study, we employed co-occurrence analysis to explore the relationships among 50 information technologies discussed in six magazines over ten years (1998-2007). Using hierarchical clustering and multidimensional scaling, we have found that the similarities of the technologies can be depicted in hierarchies and two-dimensional plots, and that similar technologies can be classified into meaningful categories. The results imply reasonable validity of our approach for understanding technology relationships and building an IT taxonomy. The methodology that we offer not only helps IT practitioners and researchers make sense of numerous technologies in the iField but also bridges two related but thus far largely separate research streams in iSchools – information management and IT management.

Keywords
Information technology management, taxonomy, co-occurrence, hierarchical clustering, multidimensional scaling

1. INTRODUCTION
The proliferation of information technologies (ITs) has been accompanied by the proliferation of information in recent decades. Opportunities emerge from such proliferation of information and technologies, making the iField an increasingly prominent and vibrant area for research and practice. At the same time, just as the explosion of information presents serious challenges in information management, the seemingly everlasting propagation of numerous ITs poses challenges in IT management. The bewildering amount of IT confronting IT practitioners and researchers renders it a challenging task to make sense of the technologies, in order to effectively manage or productively study them. In practice, IT management has been traditionally undertaken along functional lines such as hardware, software, networking, and services. Streams in IT management research, on the other hand, have mapped well onto traditional categories in practice, drawing insights from various reference disciplines such as computer science, psychology, economics, and sociology. However, recent technological and managerial advances have blurred the boundaries of traditional categories. For example, software and service have converged under the rubric of "software as a service" (SaaS). Moreover, because different types of IT may entail different cost structures, work processes, and potential returns, different ITs may require different management practices and different research methodologies. Hence, contemporary IT management practices (such as IT portfolio management) and the increasing emphasis on interdisciplinary research call for rigorous and up-to-date classifications, or *taxonomies*, of IT.

Thus far, while it has been argued that various ITs are related to varying degrees [26], it is still difficult to make sense of the relationships among technologies. For example, here is a partial list of contemporary ITs: service-oriented architecture (SOA), Web services, open source software (OSS), Web 2.0, YouTube, iPhone, blogs, and cloud computing. How are they related? How can we measure their similarities and differences? How might they be classified into meaningful categories?

IT practitioners and researchers are not well equipped to answer these questions. On the one hand, many studies in the dominant paradigm of IT management research have demonstrated that various organizational, technical, and environmental factors influence IT adoption and use [10]. As this dominant paradigm is reaching "the point of diminishing returns as a framework for supporting ground-breaking research" [10, p. 314], we note that most studies in the dominant paradigm employ single-technology research designs, leaving the relationships among ITs under-explored. On the other hand, a peripheral, yet sustained stream in IT management research has employed a variety of methods to classify technologies, practices, and/or research topics in IT [see a most recent review in 6]. Most studies in this stream had to explicitly or implicitly rely on domain experts to evaluate the similarities or differences among various technologies [e.g., 7]. Although experts may be skillful in detecting the subtleties within and across the different types of technologies, expert evaluation is often (1) biased towards the views of specific experts contributing to specific studies, (2) static in the time of such evaluations, and (3) difficult to scale up to examine the relationships among a large number of ITs. Therefore, considering the current status of the IT management literature, we raise this research question: *How can the relationships among a large number of ITs be described and analyzed in a representative, dynamic, and scalable way?*

We address this question in this study by offering a representative, dynamic, and scalable methodology to understand technology relationships and build IT taxonomies. Our approach, combining co-occurrence analysis, hierarchical clustering, and multidimensional scaling, lends itself well to automation and complements extant expert-based methods. In the following, we first briefly review the current approaches to taxonomy creation in general and IT taxonomy specifically. Then we illustrate our approach with an empirical study of 50 ITs over ten years. And finally we conclude by discussing the validity, benefits, and limitations of our approach for IT management research and practice.

2. TAXONOMY AND IT TAXONOMY

2.1 Taxonomy for Information Management

Today organizations are employing an increasing number of types of IT and an increasing amount of IT. Consequently, the amount of data collected and stored by various ITs increases exponentially. Content services, a popular technique in information management, migrate data from various sources to a common pool. Because content services do not provide inherent organizational structure for the pooled data, a taxonomy should be created to allow users to efficiently and effectively browse and use information. Hence, taxonomy creation is an important element of content management in organizations [14, 16].

A taxonomy is a classification scheme (often hierarchical) of information components (for example, terms, concepts, graphics, sounds) and their interrelationships [13]. Taxonomy creation is usually a "top-down process" by which domain experts provide an overview of the domain, list categories and features of each category, and finally classify categories into broader classes according to how similar the features of the categories are [17]. Categories that do not mach current classes are put aside until enough categories with sufficiently similar features appear to justify the creation of new classes [16]. It has been recommended that analysts use and customize pre-populated taxonomies whenever available [14].

2.2 IT Taxonomy

Sustained, despite relatively peripheral, efforts have been made in IT research to classify technologies, applications, and research topics and methods [e.g., 1, 2, 6, 7, 9, 11, 23, 25]. The usual process that produces various classification schemes and taxonomies in IT is very similar to the taxonomy creation approach for information/content management described above, except that the initial list of categories that constitute the taxonomies often comes from empirical surveys of the IT discourse. For example, three classification studies of information systems research collected keywords of publications as initial input to the taxonomy creation process [1, 2, 6]. Once the initial list has been compiled, experts (often the study's authors) arrange the items on the list according to their assessment of various features of the items. For example, Ein-Dor and Segev [7] surveyed the definitions of 17 technologies in the IT discourse, identified from the definitions 31 attributes and 27 functions, and then described the technologies by two bit-vectors: a vector of attributes and a vector of functions. Furthermore, they performed quantitative methods such as multidimensional scaling (MDS) to visualize the relationships among the technologies in terms of their relative similarity.

2.3 Limitations of Extant Approaches

To varying degrees, extant methods for creating taxonomies in general and specifically for IT rely on experts. While expert opinions are valuable in grounding the taxonomy in specific domains and detecting subtleties in the relationships among categories, current approaches have several limitations.

First, the structures of extant taxonomies represent a relatively narrow set of views from only a few experts. For instance, the choice of features (such as attributes and functions of IT) for classification depends on the specific opinions or background knowledge of the experts who participate in the study. Second, taxonomies built by this approach seem static, fixed at the time when experts created them. Efforts to update existing taxonomies are few and far in between. For example, the ACM Computing Classification System currently being used was created in 1998. As another example, the official Keyword Classification Scheme for Information System Research was last updated in 1993 [2]. Finally, such scant efforts to update existing taxonomies may be due to another limitation – methods relying on experts are not scalable, lending themselves poorly to automation. As the number of ITs increases, the effort by human experts to describe each technology according to its attributes and functions increases, and the reliability of that classification work may decrease. Addressing these limitations of the extant approaches to IT taxonomy creation, in this study we develop a methodology that builds upon the existing methods of analyzing IT discourse, but allows wider representations of opinions, dynamic updating at multiple points of times, and large-scale automated analysis of a large number of technologies.

3. DEVELOPING A REPRESENTATIVE, DYNAMIC, AND SCALABLE APPROACH

In this section, we describe our approach with an empirical study as an illustration.

3.1 Data Collection

There are many outlets for the IT discourse, including books, magazines, conferences, blogs, wikis, and many others, where discourse data may be collected. In order to illustrate how our methodology works in IT management research, we decided to focus on two IT trade magazines (*InformationWeek* and *Computerworld*), two business magazines (*BusinessWeek* and *The Economist*), and two news magazines (*Newsweek* and *US News & World Report*). As described below, the scale of the data we collected from the six magazines is large enough for us to demonstrate the scalability of our approach. In addition to the scale, our data is also diverse, representing a wide range of views on IT and broader topics.

We downloaded from the Lexis/Nexis online database all articles published during a ten-year period (1998-2007) in the six magazines, totaling about 220,000 articles. Meanwhile, we compiled a list of 50 IT concepts (Table 1), ranging from enterprise software (e.g., CRM) to personal gadgets (e.g., iPod), from abstract concepts (e.g., artificial intelligence) to concrete products/services (e.g., YouTube), and from highly popular (e.g., e-business) to less well-known concepts (e.g., digital subscriber line – DSL). Admittedly, this list is *ad hoc*, but it serves the illustration purpose because the list covers a broad range of technologies in the examination period. We then extracted from the articles all paragraphs that contain any of the technologies on the list. In doing so, we considered multiple possible labels for each technologies, plural forms, and acronyms unique to the technology. For example, in extracting paragraphs containing "digital subscriber line," we also included paragraphs mentioning "digital subscriber lines" and "DSL." In total, 105,400 paragraphs containing at least one technology on the list were extracted from the full text of the articles published the six magazines.

Table 1. Information technologies examined in the study

Label	Full Name of Information Technologies
AI	Artificial intelligence
ASP	Application service provider
BI	Business intelligence
Blog	Blog
Bluetooth	Bluetooth
BizProReen	Business process reengineering
CloudCom	Cloud computing
CRM	Customer relationship management
DigiCam	Digital camera
DLearn	Distance learning
DSL	Digital subscriber line
DecisionSS	Decision support system
DW	Data warehouse
eBiz	Electronic business
eCom	Electronic commerce
EDI	Electronic data interchange
ERP	Enterprise resource planning
ExpertSys	Expert system
GPS	Global positioning system
Grpware	Groupware
IM	Instant messaging
iPhone	iPhone
iPod	iPod
KM	Knowledge management
Linux	Linux
Multimedia	Multimedia
MP3	MP3 player
MySpace	MySpace
NeuralNet	Neural net
OLAP	Online analytical processing
OSS	Open source software
Outsource	Outsourcing
PDA	Personal digital assistant
RFID	Radio frequency identification
SmartCard	Smart card
SCM	Supply chain management
SFA	Salesforce automation
SocNet	Social networking
SOA	Service oriented architecture
Telecommute	Telecommuting
TabletPC	Tablet PC
UtiComp	Utility computing
Virtualization	Virtualization
VPN	Virtual private network
Web2	Web 2.0
WebServ	Web services
WiFi	Wi-Fi
Wiki	Wiki
Wikipedia	Wikipedia
YouTube	YouTube

3.2 Data Analysis

To make sense of the relationships among the technologies, we focused on the initial step of exploring the similarity of the technologies. One approach is to automatically infer similarity of technologies from their co-occurrences in the same unit of discourse (e.g., an article, paragraph, or sentence) [22]. We also used hierarchical clustering analysis and multidimensional scaling to classify the technologies.

3.2.1 Co-occurrence

Co-occurrence of words or terms has been used in various fields such as computational linguistics [3, 4] and information retrieval [21] to study the relationships among words or terms. For example, Spence and Owens [22] used co-occurrence to evaluate the strength of word association. They found that related pairs of nouns co-occur considerably more often than unrelated pairs. Their finding suggests that co-occurrence frequency may indicate the strength of word association.

Analysis of co-occurrence should define a proper size of the window where words or terms co-occur. A window size can be a certain number of words or characters [e.g., a window of 250 characters in 22] or a logical division of an input text [19]. We chose paragraph as the window size because it sufficiently captures the context for describing related technologies.

To measure co-occurrence at the paragraph level, from the 105,400 paragraphs we initially extracted, we selected paragraphs containing two or more ITs in Table 1. This filtering process returned approximately 12,000 paragraphs. Then we constructed a 50x50 co-occurrence matrix with each row or column representing a technology on the list. The value in each cell of the matrix represents the number of paragraphs containing the respective pair of technologies. This co-occurrence matrix is a matrix of similarity. In order to perform subsequent classification and visualization techniques that are based on dissimilarity

measures, we transformed the similarity matrix to a dissimilarity matrix with the formula: *1/(x+0.1)*.

3.2.2 Hierarchical clustering

Cluster analysis is the process of grouping objects into unknown clusters such that the within-group variation is minimized and the between-group variation maximized [8]. The agglomerative hierarchical clustering method groups objects on a series of levels, from the finest partition, in which each individual object forms its own cluster, and successively combines smaller clusters into larger ones until all objects are in one cluster. Agglomerative hierarchical clustering employs an aggregation criterion, or "linkage rule," to determine how the distance between two clusters should be calculated based on the distance scores of pairs of objects. The most well-known aggregation criteria are single link, complete link, and average link [12]. The distance between two clusters is represented by the minimum, maximum, or average distance between any pair of objects, one object from each cluster. In single link clustering, two clusters with the smallest minimum pairwise distance are merged in each step. In complete link clustering, two clusters with the smallest maximum pairwise distance are merged in each step. And average link clustering is a compromise between the other two methods. We used the average link in this study because of its robustness [5].

3.2.3 Multidimensional scaling

Previous research has found that applying multidimensional scaling (MDS) and clustering separately to the same proximity data results in greater insight into the structure underlying the data and can detect more subtle and complex relationships than either method used alone [15, 18, 20]. Both clustering and MDS are visualization techniques. The key difference between the two techniques is that MDS provides a spatial representation of the data, while clustering provides a tree representation [15].

Based upon a matrix of item-item similarities or dissimilarities, an MDS algorithm assigns a location to each item in a space such that the distances between the items correspond as closely as possible to the measured dissimilarities between the items. In other words, the proximity of items to each other in the space indicates how similar they are. We used the MDS procedure based on the ALSCAL or alternating least squares scaling [24], a popular MDS algorithm. For easy interpretation of the result, we chose to present the MDS solutions in two-dimensional scatter plots.

4. RESULTS

Our clustering analysis of the transformed co-occurrence matrix generated a hierarchical structure of 50 technologies in a dendrogram (Figure 1), where vertical lines show joined clusters and the position of the lines on the scale from 1 to 25 indicates the distance at which clusters are merged. By inspecting the dendrogram, we have identified eight clusters, all of which merged around 5 in the 25-point scale. These eight clusters are indicated by the intersections between the dendrogram and the vertical dotted line in Figure 1. Table 2 summarizes the membership of each cluster. In Figure 2, we depict the 50 ITs in a two-dimensional MDS plot. Following Shepard and Arabie's [20] suggestion, we have used different colors to represent the eight clusters identified in the clustering analysis. Generally speaking, most of the technologies in the same cluster are located close to each other in the MDS plot. We describe several clusters in more details below.

Table 2. Membership of the clusters

Cluster	Labels of Information Technologies*
1	eBiz, eCom, CRM, ERP, Outsource, ASP, SCM, SFA, EDI, Grpware, KM, BizProReen (BPR)
2	RFID, SmartCard
3	BI, DW, OLAP, DecisionSS
4	AI, NeuralNet, ExpertSys
5	DSL, VPN, Telecommute, DLearn
6	Bluetooth, WiFi, PDA, GPS, iPod, MP3, DigiCam, Multimedia, iPhone, TabletPC
7	Wiki, Wikipedia, MySpace, SocNet, Blog, YouTube, Web2.0, IM
8	UtiComp, Virtualization, Linux, OSS, SOA, WebServ, CloudCom

***** Please see the full names of the IT labels in Table 1.

Cluster 1 includes twelve IT concepts. All of them are enterprise IT applications except outsourcing, which is a strategy for managing enterprise IT. Business process reengineering (BPR) was the last to join the cluster, suggesting that it is the least similar to the others in the cluster. This situation may explain why BPR looks like an outlier in the cluster in the MDS plot (Figure 2). Cluster 5 includes four IT concepts. Among them, digital subscriber line and virtual private network are both telecommunication technologies, which may be employed in the other two IT applications (telecommuting and distance learning). Cluster 6 has ten IT concepts, all related to mobile or wireless technologies. Some, such as bluetooth and Wi-Fi, are the underlying mobile technologies. Others, such as TabletPC and PDA, are the devices enabled by the wireless/mobile technologies. Cluster 7 has eight IT concepts. They are the so-called Web 2.0 technologies that have become highly popular in recent years. Lastly, Cluster 8 includes seven IT concepts of similar type such as utility computing, Web service, and cloud computing.

According to the agglomeration schedule, a series of steps during clustering, we were able to identify twelve pairs of ITs considered most similar to each other in the list (Table 3). The pairs include, for example, e-business and e-commerce, iPod and MP3, and artificial intelligence and neural net. These pairs are compatible with even rudimentary understanding of these technologies.

Table 3. Pairs of most similar ITs

Pair	IT*	Pair	IT*
1	eBiz, eCom	7	Bluetooth, WiFi
2	CRM, ERP	8	iPod, MP3
3	Linux, OSS	9	DSL, VPN
4	BI, DW	10	Grpware, KM
5	SOA, WebServ	11	AI, NeuralNet

6	MySpace, SocNet	12	Wiki, Wikipedia

* Please see the full names of the IT labels in Table 1.

5. DISCUSSION

5.1 Validity of the Approach

The results illustrate that co-occurrence data can be utilized for classification. Our co-occurrence analysis, supplemented by the two visualization techniques, has yielded results that can be interpreted fairly easily, even without the presence of sophisticated expert knowledge of the various domains that our list covers. The face validity we have seen in this illustration study gives us reasonable confidence in applying our methodology to other circumstances where *a priori* knowledge is unavailable, such as the cases of new or unknown technologies.

5.2 Benefits of the Approach

Our approach has several advantages. First, this approach is more representative than extant methods for taxonomy creation, which often rely on a small number of experts and represent a narrow set of views. The community of organizational and individual stakeholders represented by any of the magazines we selected to study in this project is obviously larger than any group of experts enlisted in previous classification studies. Out of curiosity, we sorted the data by each magazine and performed the same analysis. Figure 3 compares the dendrograms we produced using the *InformationWeek* and *BusinessWeek* data. The differences in the IT hierarchies signify the different structures of IT knowledge that were developed in the two communities. In the illustration above, we pooled the data from all six magazines, making the results even more representative of the broader socio-technical context in which technological innovations emerge and evolve.

Second, speaking of evolution, we note that our approach allows updating taxonomies at multiple points of time, enabling longitudinal analysis of the dynamic relationships among technologies. In fact, technologies do change over time and their relationships change too. For illustration, we divided the *InformationWeek* data into two five-year periods (1998-2002 and 2003-2007) and performed the same analysis on the two subsets of data. Figure 4 shows the dendrograms for the two periods. One notable difference between the two dendrograms is that e-business and e-commerce, almost interchangeable in the first period, diverged in the second period.

Lastly, this approach is scalable. The study has examined the six magazines for 50 ITs over ten years, already surpassing the scale and scope of many IT classification studies. While we have used six magazines for this illustration, automation in this approach is not limited in the type or number of discourse outlets or the type or number of technologies.

5.3 Limitations and Future Research

The benefits we just discussed can be realized only within the limitations of this approach. First, the 50-IT list, despite the diversity in it, is an *ad hoc* list that we generated based on our own knowledge of the various domains in IT. Future research should develop a more systematic way to identify technology categories to be included in a taxonomy. While it is never our intention to exclude human knowledge from the selection process, we suggest using automated topic detection techniques to generate a preliminary list and then developing criteria for selection by humans. Second, the quality of the taxonomy must be assessed against "ground truth," which is currently absent in our approach. Therefore, a logical next step is to search for or develop baselines for evaluating quality. Lastly, the usefulness of a taxonomy will ultimately be determined by how well it satisfies users' requirements, which vary across user groups such as IT managers and IT researchers. Consequently, future research should collect requirements from target user groups, build taxonomies according to specific requirements, and test usability in different user groups.

6. CONCLUSION

In conclusion, our combined use of co-occurrence analysis, hierarchical clustering, and multidimensional scaling has given rise to a representative, dynamic, and scalable approach to building IT taxonomies. Properly developed taxonomies are useful in many aspects of IT management. For providers of IT products and services, a taxonomy empirically developed may complement the product categories designated in a top-down design process. For adopters of IT products and services, taxonomies are needed for portfolio management. For scholars in the iField, our approach not only helps many make sense of the complex and dynamic relationships among numerous technologies, but also bridges two related, but thus far largely separate streams of research in iSchools: information management and IT management. As we have shown, commonplace information management techniques such as co-occurrence analysis and clustering can be profitably integrated and applied to solve problems in IT management. Hence the moral of this study: There is a large amount of information about a large amount of IT. A large amount of IT generates a large amount of information. Therefore, effective IT management and effective information management take place hand in hand.

7. ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grants No. IIS-0729459 and SBE-0915645.

8. REFERENCES

[1] Barki, H. and Rivard, S. 1988. An information systems keyword classification scheme. MIS Quarterly. 12, 2, 299-322.

[2] Barki, H., Rivard, S., and Talbot, J. 1993. A keyword classification scheme for is research literature: An update. MIS Quarterly. 17, 2, 209-226.

[3] Burgess, C. and Lund, K. 1997. Parsing constraints and high-dimensional semantic space. Language and Cognitive Processes. 12, 177-210.

[4] Burgess, C. and Lund, K. 1997. Representing abstract words and emotional connotation in high-dimensional memory space. In Proceedings of the 19th Annual Conference of the Cognitive Science Society (Mahwah, NJ). Lawrence Erlbaum Associates. Inc., 61-66.

[5] Chandon, J.-L. and Pinson, S. 1981 Analyse typologique: Théories et applications. Masson.

[6] Dwivedi, Y., Mustafee, N., Williams, M. D., and Lal, B. 2009. Classification of information systems research revisited: A keyword analysis approach. In Proceedings of Pacific Asia Conference on Information Systems (Hyderabad, India).

[7] Ein-Dor, P. and Segev, E. 1993. A classification of information systems: Analysis and interpretation. Information Systems Research. 4, 2, 166-204.

[8] Everitt, B., Landau, S., and Leese, M. 2001 Cluster analysis. Oxford University Press.

[9] Farbey, B., Land, F. F., and Targett, D. 1995. A taxonomy of information systems applications: The benefits' evaluation ladder. European Journal of Information Systems. 4, 1, 41-50.

[10] Fichman, R. G. 2004. Going beyond the dominant paradigm for information technology innovation research: Emerging concepts and methods. Journal of the Association for Information Systems. 5, 8, 314-355.

[11] Fiedler, K. D., Grover, V., and Teng, J. T. C. 1996. An empirically derived taxonomy of information technology structure and its relationship to organizational structure. Journal of Management Information Systems. 13, 1, 9-34.

[12] Hansen, P. and Jaumard, B. 1997. Cluster analysis and mathematical programming. Mathematical programming. 79, 1-3, 191-215.

[13] Harris, K., Caldwell, F., Linden, A., Knox, R., and Logan, D., "Taxonomy creation: Bringing order to complexity," Gartner, Inc. QA-20-8719, 10 September 2003.

[14] Jagerman, E. J. 2006 Creating, maintaining and applying quality taxonomies. Lulu.com.

[15] Kruskal, J. B. 1977. The relationship between multidimensional scaling and clustering. In Classification and clustering, J. Van Ryzin, Ed. Academic Press, New York, 17-44.

[16] Lambe, P. 2007 Organising knowledge: Taxonomies, knowledge and organisational effectiveness. Oxford: Chandos.

[17] Logan, D., "Best practices for taxonomy creation," Gartner, Inc. G00167683, 2009.

[18] Napior, D. 1972. Nonmetric multidimensional techniques for summated ratings. In Multidimensional scaling; theory and applications in the behavioral sciences, R. N. Shepard, et al., Eds. Seminar Press, New York, 157-178.

[19] Schvaneveldt, R. W. 1990 Pathfinder associative networks: Studies in knowledge organizations. Ablex Pub. Corp.

[20] Shepard, R. N. and Arabie, P. 1979. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. Psychological Review. 86, 2, 87-123.

[21] Smadja, F. 1993. Retrieving collocations from text: Xtract. Computational Linguistics. 19, 143-177.

[22] Spence, D. P. and Owens, K. C. 1990. Lexical co-occurrence and association strength. Journal of Psycholinguistic Research. 19, 5, 317-330.

[23] Swanson, E. B. and Ramiller, N. C. 1993. Information systems research thematics: Submissions to a new journal, 1987-1992. Information Systems Research. 4, 4, 299-330.

[24] Takane, Y., Young, F. W., and De Leeuw, J. 1977. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. Psychometrika. 42, 1, 7-67.

[25] Vessey, I., Ramesh, V., and Glass, R. L. 2005. A unified classification system for research in the computing disciplines. Information and Software Technology. 47, 4, 245-255.

[26] Wang, P. 2009. Popular concepts beyond organizations: Exploring new dimensions of information technology innovations. Journal of the Association for Information Systems. 10, 1, 1-30.

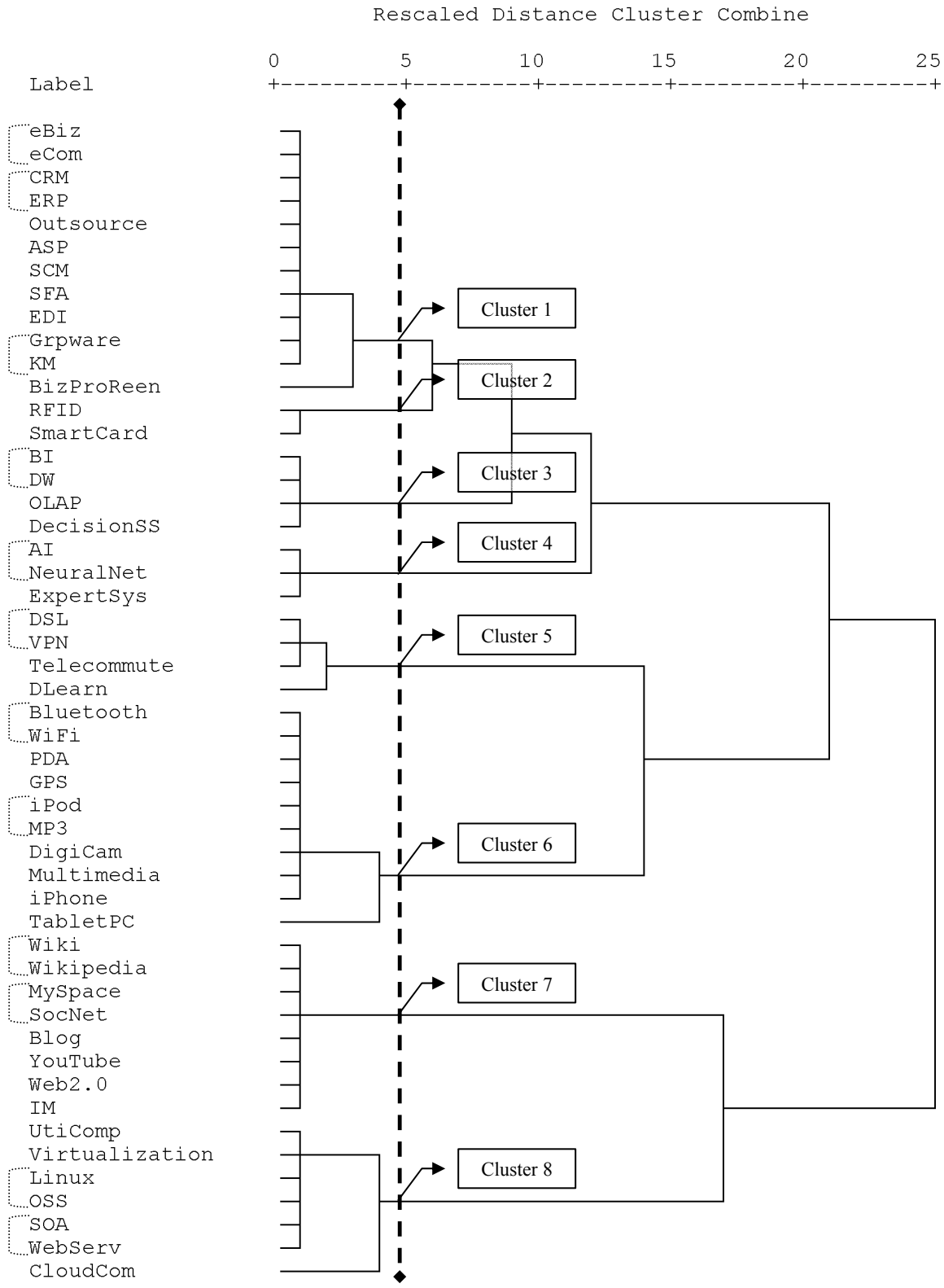


Figure 1. Dendrogram generated from hierarchical clustering analysis

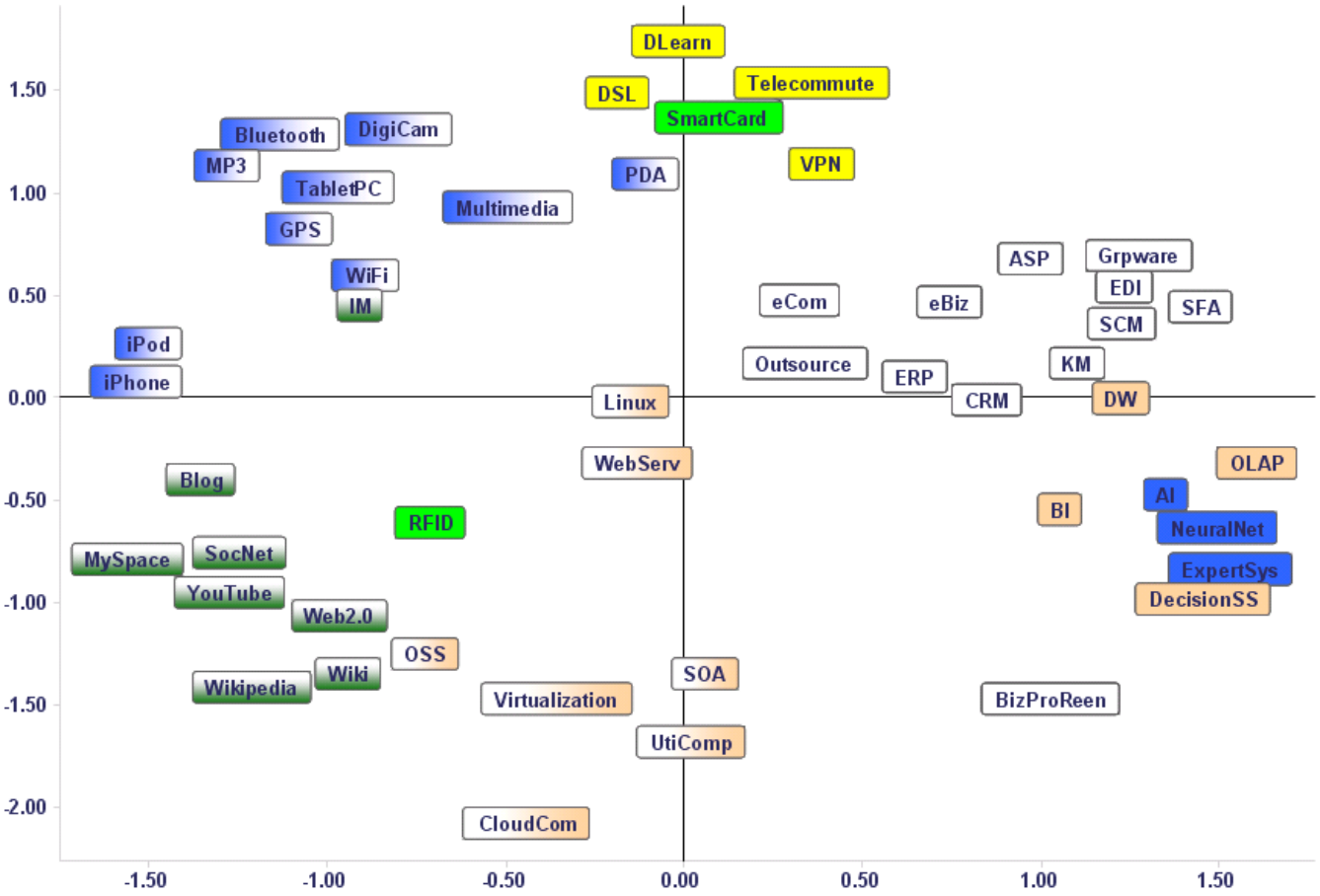
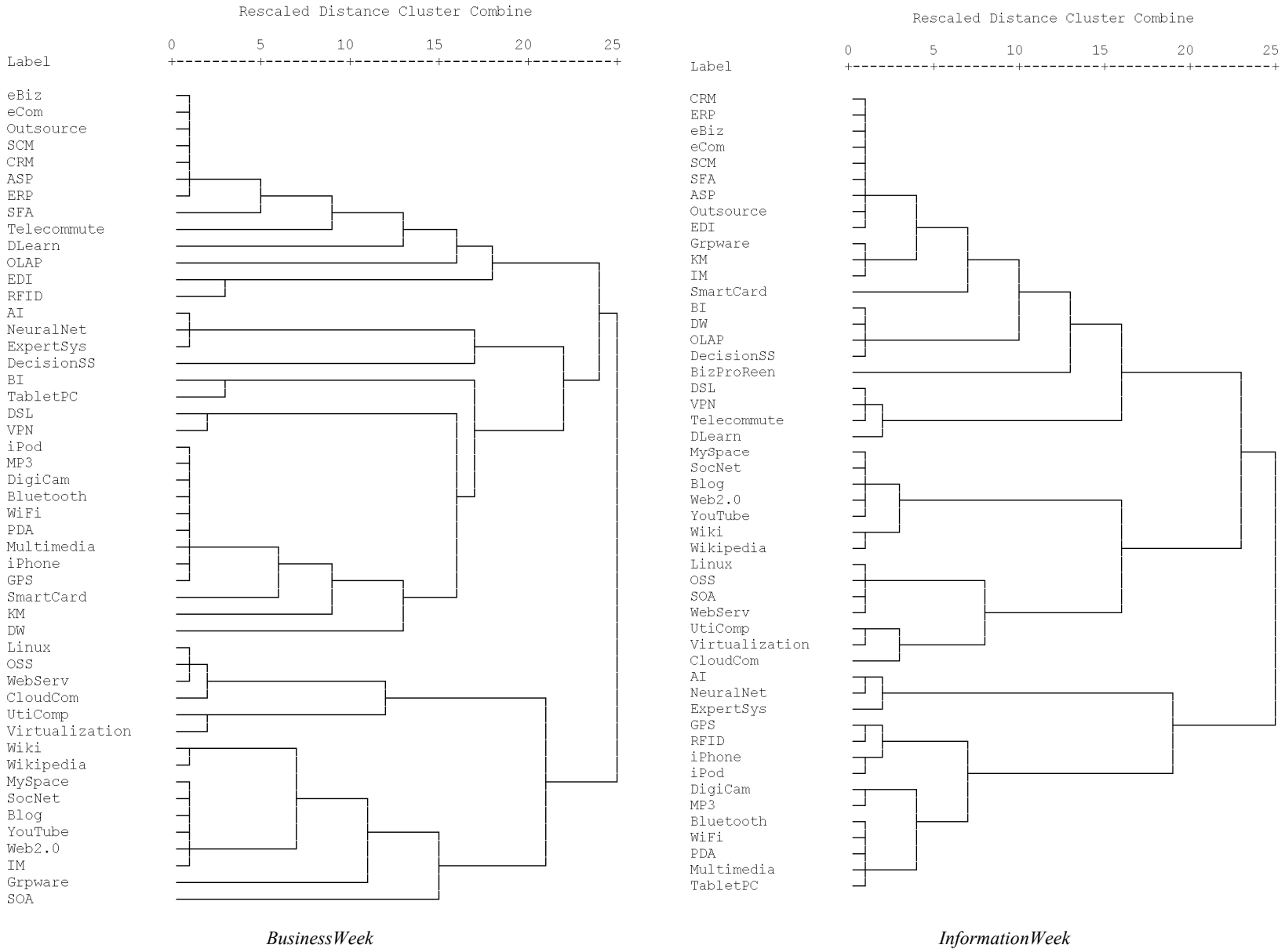


Figure 2. A multidimensional scaling (MDS) plot of the 50 ITs (six magazines, 1998-2007)



BusinessWeek *InformationWeek*

Figure 3. Dendrograms generated from hierarchical clustering analysis of data in individual magazines (1998-2007)

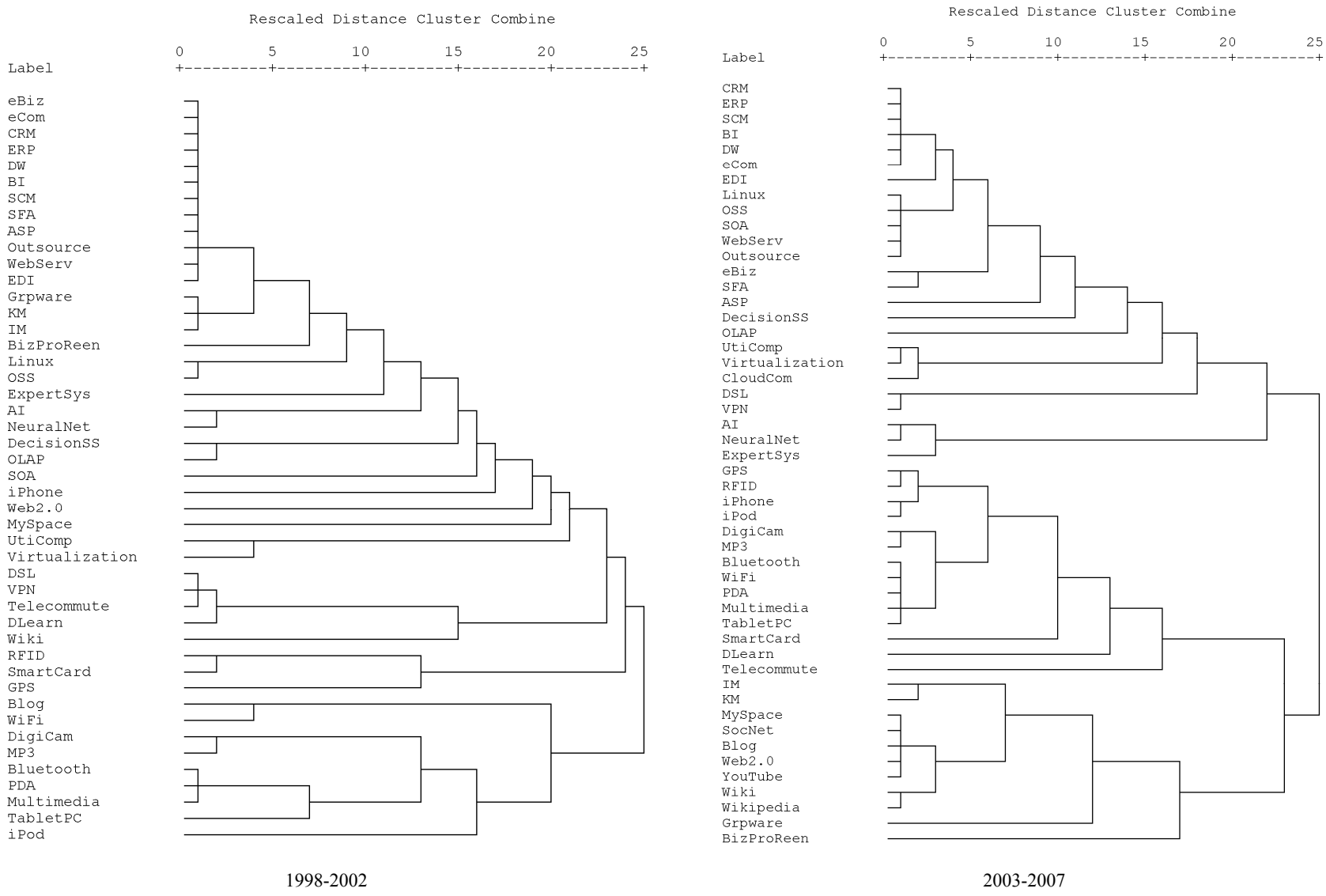


Figure 4. Dendrograms generated from hierarchical clustering analysis of *InformationWeek* data in 5-year periods

Cultural Heritage Information Dashboards

Richard J. Urban
Graduate School of Library and Information Science
University of Illinois
Champaign, Il
rjurban@illinois.edu

Piotr Adamczyk
Metropolitan Museum of Art
1000 Fifth Ave.
New York, New York
piotr.adamczyk@metmuseum.org

Michael B. Twidale
Graduate School of Library & Information Science
University of Illinois
Champaign, Il
twidale@illinois.edu

ABSTRACT

Large-scale aggregations of digital collections from libraries, archives and museums offer users unprecedented access to cultural heritage materials. But they also have failed to incorporate important contextual information that allows users to develop an understanding of the significant features of purpose-built collections. This paper explores the development of information dashboard prototypes that provide users a high-level overview of cultural heritage collections. Two case studies using rapid-prototyping methodologies are presented.

Categories and Subject Descriptors

D 2.1 Requirements/Specifications [Elicitation methods]

General Terms

Design, Human Factors

Keywords

Rapid prototyping, information visualization, dashboard, cultural heritage

1. Introduction

Currently available digital collections from libraries, archives and museums represent just the tip of the iceberg – millions more analog resources could potentially be added in the next decade. As these resources move from physical repositories into networked environments they run the risk of losing connections to institutional and curatorial contexts that make "collections" significant as curated, identifiable wholes. Traditional approaches to description have relied on textual metadata augmented by the tacit knowledge of the professionals who care for collections. In order to gain access to physical collections it was necessary to work through local institutional infrastructures [[1], [2]]. Although institutional websites may recapture some of the essence of this infrastructure, many of the subtle affordances of physical modes of access have not carried forward into the digital realm.

While traditional modes of collection access may have been sufficient for distributed institution based materials, they also suffered from the known problem of "collections understanding" among both professionals and users. Lee discovered that the meaning of "collection" is difficult to pin down even among librarians, archivists and museum professionals who work with them every day. Professional perspectives of how collections are managed diverge from the user's perspective of how collections are used, but it is the professional perspective that has largely shaped collection-level descriptive practices – often without a clear grasp of the role that collections play in information seeking behaviors [3]. Professional practices for collection-level and item-level metadata have also proceeded independently of each other without consideration of how thy could be mutually supportive. The problems of "collection understanding" are further exacerbated by the rapid changes that online access to cultural heritage materials brings. Large-scale aggregations containing collections from multiple institutions add to the problem and may make individual items "informationally small" due to the lack of collection-level metadata [4]. Many current descriptive practices are based on an assumption that the institution is the gateway. However usage logs show that users increasingly arrive at cultural heritage materials not through the front door, but through broader information seeking behaviors on the open web that drop them into the middle of a collection or aggregation. In such cases, users may not be able to discern who owns a collection, how big it is, what it contains, its relative strengths compared to other collections and the significance of its constituent items.

Fortunately, lower barriers to information visualization are creating opportunities to re-imagine how we represent the contours of collections. Our research seeks to restore the contextual information that digital collections have lost through the use of "collection dashboards." Using a user-centered rapid-prototyping method known as "patchwork prototyping" we elicited initial concepts from humanities scholars and library, archive and museum professionals in conference settings. These concepts served as the basis for two prototypes - one which uses OAI-PMH Dublin Core metadata aggregated by the IMLS Digital Collections and Content (IMLS DCC) Project and the other using publicly available data about the Metropolitan Museum of Art's collections. In both cases, the quality of metadata syntax and semantics created significant barriers to successfully automating the generation of dashboards. We discuss a number of issues raised by the initial prototypes.

2. Background

2.1 Visualizing Cultural Heritage Collections

As the power of graphics processing has increased, the ability to visualize larger and more complex data sets increasingly has come within the reach of the cultural heritage sector. This lower threshold, combined with increasing amounts of metadata about cultural heritage materials, creates an opportunity to use exploratory information visualizations to support information seeking activities. The research literature on information visualization is both broad and deep, but has largely been confined to the lab, beyond the reach of many libraries, archives and museums. The work of Martin Wattenberg, Fernanda Viégas

and others interested in "democratizing" information visualization has made these techniques publicly available through web services such as Many Eyes™ and the Google Visualization API [5]. Visualization tools have left the lab and entered the mainstream through interactive sites such as the New York Times Visualization Lab, representations of personal social networks or previously incomprehensible government data. [6], [7], [8].

Like many of these examples, libraries, archives and museums posses a treasure trove of complex data describing their collections. Traditionally this information is put in the service of text-based search and retrieval systems, but these new and more broadly available information visualization techniques offer the opportunity to re-imagine how this information could be used or improved to facilitate such visualizations. New publicly available visualization approaches can benefit from the foundational research in visualizing cultural heritage collections. Green, et al. developed a novel browser for Library of Congress American Memory collections that included timelines, topical browsers and geospatial views of collections [9]. Also using the American Memory collections, Derthick's *Bungee View* presents a faceted browsing interface that also highlights the relative size of facets in a collection [10]. Using archival descriptions, the ArchivesZ project and the *Visible Archives* attempt to provide users a sense of the relative size and scale of archival series [11], [12].

A related area is the development of "information dashboards" that may include multiple information visualizations. The dashboard metaphor has its roots in the business sector where they are used to track large, complex and changing data from financial markets, industrial processes or inventory flow [13]. A few libraries have adopted the dashboard metaphor to make circulation and reference statistics available to managers – or the public, such as George Legrady's *Invisible Made Visible* installation at the Seattle Public Library [14]. Dashboards have also made inroads to the cultural heritage sector through financial products that allow non-profits to view endowments and investments. The Indianapolis Museum of Art (IMA) has adopted the dashboard metaphor to provide visitors with information about attendance statistics, power consumption and what is blooming in the garden this week [16]. Part of the aim of the IMA's dashboard is to make this information more public – a kind of "radical transparency" whereby an organization chooses to share data as a means to include a wider constituency in its core processes [17]. For a museum such as the IMA, this can be a kind of public outreach or recruitment. The data to be shared may not on its own be very exciting, but when combined and visualized it can tell a compelling story about the contributions of an institution to a community. This exploration of collections dashboards seeks to extend this kind of transparency to the collections themselves.

2.2 Patchwork Prototyping

In order to explore the problem space of collections dashboards, this research adapted the patchwork prototyping approach [17]. This rapid-prototyping method combines user-centered design with high-fidelity prototypes constructed using open-source software and freely available web services. While this "mashup" approach has been used to generate interesting web services (such as a Google Map of Craigslist apartment listings), patchwork prototyping explicitly ties them to prototyping tasks. Because it relies on "off-the-shelf" components loosely stitched together, patchwork prototyping can fills the gap between low and high fidelity prototypes.

By itself, patchwork prototyping does not address the common problem of access to intended users of novel products. The number of people willing and able to help may be limited, especially if a certain level of skill is required, and the time they can devote to helping is also limited. In the case of this project, our target users are professional historians, researchers and skilled amateurs. Additionally librarians, archivists and museum professionals because of their intermediary experiences can also inform our design. Along with this informant constraint common to many projects, we wanted to try to generate and exploit new opportunities for informing design. How might target groups contribute in very fast, lightweight, low commitment ways, other than more conventional participatory design meetings that typically take at least an hour? One idea was to do some of our design work at conferences. Many conferences have a demo session where a working application or prototype is shown off. Attendees typically offer informal comments and although these can and do inform the designers, this does not seem to be considered as a design process. What if instead of doing a conventional demo of a finished product, the demonstration session was used as a platform for participatory design?

When demonstrations are used as a user-centered-design (UCD) process rather than a conventional demo, it can have very desirable properties. Demonstrations clearly simplify recruitment. The process of participating / volunteering is less effortful, one might even say less aggressive than asking a person to help in a conventional UCD session. Interactions are understood to be typically short, typically 5-10 minutes, lowering the perceived time commitment. Participation can be incremental, starting from a peripheral observation and calling out questions or comments, gradually moving up to hands-on interaction. It is contextually acceptable to participate for a short while and then move on to another demo. A demo session is noisy and messy, with considerable movement affording multiple conversations and many kinds of interactions. Finally it is very easy for a participant to politely disengage and go to another booth whenever they want to. Ease of disengagement helps in encouraging initial engagement because the level of implied commitment is less.

3. Case Studies

3.1 Opening History

The Institute for Museum and Library Services Digital Collections and Content (IMLS-DCC) Project began in 2002 in order to aggregate collection-level and item-level descriptions from digitization projects funded by the Institute of Museum and Library Services (IMLS). Beginning in 2007, the IMLS DCC project expanded the scope of its aggregation activates to digital collections related to American history. Known as Opening History this aggregation currently contains more than 770 collection-level descriptions and over 950,000 item-level descriptions from more than 300 different libraries, archives and museums (LAMs) [18]. Because of the diversity of materials aggregated, we have had a keen interest in the role that collections play in users' ability to identify useful resources [[19], [20]. An important part of our current workplan is exploring how to better represent the content and contexts of Opening History for scholarly use and users.

Previous development work for IMLS DCC/Opening History has followed a traditional approach of using low-fidelity paper prototyping to generate design ideas along with high-fidelity working prototypes that were used for usability testing. While the

low-fidelity approaches were useful in generating ideas, they were unable to capture the dynamic interactions that a scholar encounters when working with rich collections of cultural resources. Where such a scholar is not particularly technically sophisticated, paper prototypes can be very engaging in encouraging discussions about design. They can help in considering a static presentation of a certain set of results, but can prove difficult in supporting the envisaging of a dynamic interaction. By contrast, high-fidelity prototypes do support consideration of interaction but are slow to develop and thus difficult to iterate within the available schedule, leading to design lock-in.

Desirable visualizations for the Opening History Collection Dashboard were generated through interaction with approximately one hundred library & information science researchers, digital humanities scholars, librarians, archivists and museum technology professionals who attended five different conferences. Two of these design exercises took place took place within l conference exhibit booths. Participants in the sessions were provided with paper templates of visualizations extracted from research papers, known visualization projects and examples from other library, archives and museum interfaces. These cut-out visualizations had small magnets attached so that they could be easily added to or moved around on a light metal board. Although these visualizations represented various other kinds of data, participants were also asked to add annotations that suggested modifications or particular use cases for cultural heritage collection, including drawing desired visualizations on sticky notes.

While some participants were intrigued by more novel visualizations (such as one which showed the physical scale of different sized objects), most indicated a strong preference for visualizations that addressed the traditional questions of who, what, where, and when [21]. Participants were encouraged to do ego-centric design by suggestions like "can you put the items together to make an interface that you personally might find useful?" The separate discussions of what was selected, what was rejected and what was missing were then used to inform a design for less idiosyncratic use. Many participants were familiar with similar kinds of configurable component displays such as in iGoogle and Yahoo! although there is very little like it for online access to digitized collections. Participants noted the desirability of personalization, describing which kinds of elements they would find useful. However this was not just a matter of per-person customization, but also per-use-type. Several people described different activities that they did that would benefit from different combinations of components.

Using input from participants in our demonstrations, we developed an initial prototype that provides a minimal level of functionality for visualizations of Spatial Coverage (Where), Date Items Created (When), Item Types in Collection (What), and a limited list of the top 50 Subjects (What) based on the IMLS DCC collection-level OAI-PMH metadata. Value frequencies were generated using the SIMILE Gadget utility and converted into a comma-separated-value (CSV) tabular format. These individual value/frequencies were then visualized using the Google Visualization API or Many Eyes™ services. The IMA Museum Dashboard, a Drupal template module, was used to stitch together individual visualizations into a complete dashboard . [22]

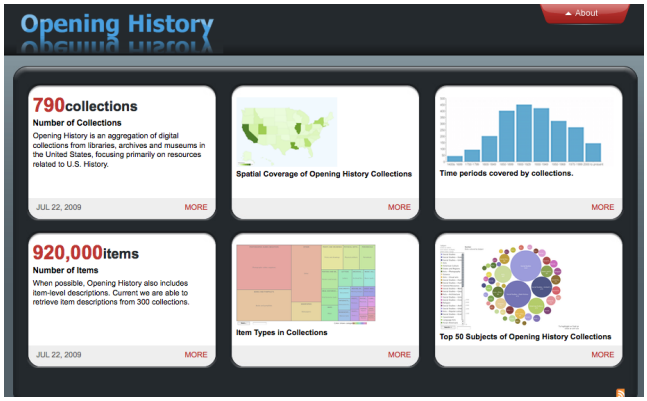


Figure 1 Opening History Dashboard Prototype

While this prototype provides an overall view of the IMLS DCC aggregation (based on collection-level metadata), it does not yet provide a similar view for each of the collections with item-level metadata aggregated by IMLS DCC. As discussed below, the heterogeneity and quality of item-level metadata posed significant challenges to quickly generating the data necessary to provide accurate visualizations

3.2 Metropolitan Museum of Art

The Metropolitan Museum of Art collections include more than two million works of art from all over the globe spanning five thousand years of history. Information about the collection is presented online in a straightforward manner – faceted by department, searchable by keyword, and sortable by individual fields. The only aggregate collection level information presented is the number of objects in a department. Individual object pages provide an image of the object along with the object description and attendant metadata, but no efforts have been made to further contextualize the object within the collection.

Inspired by other efforts at making cultural heritage information more transparent, the Museum Pipes project started as an examination of the character and quality of the results available through web services retrieved when passing the most minimal of metadata, e.g. artist name, date, or a single keyword [23]. To date, Museum Pipes dashboards have used both Curatorial Department and Museum Library datasets. These early prototypes have been intended exclusively for use by museum professionals. Though subsequent and more thoroughly designed versions of these dashboards could eventually serve the needs of a broader audience, in their current state they are best suited to more specialized tasks. In particular recognizing patterns and exceptions in item level metadata, made more recognizable by their presentation in aggregate, suggesting which collections may require more detailed scrutiny before online publication.

The most robust object level tool from Museum Pipes is an information aggregator that, when provided with a Metropolitan Museum online unique object identifier, scrapes metadata from the Object Record View page on metmuseum.org and passes this to a wide array of third-party information repositories. The current implementation collects press accounts (from NPR, the New York Times, and the Guardian), related images (from Flickr, TinEye, and Google Images), books (from OCLC/WorldCat, Internet Archive, OpenLibrary, and Google Books), additional metadata (tags from OCLC identity records, Wikipedia/DBpedia data, and keywords extracted from the object description by other web

services), along with additional social web content (from delicious, twitter, and YouTube).

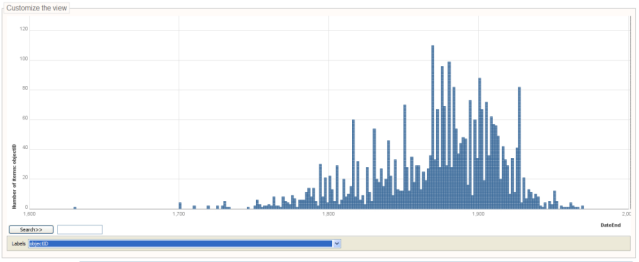


Figure 2 When were works in the American Paintings and Sculpture Collection Made?

4. Discussion

Even within these two case studies, a number of different issues arose – either challenging our ability to develop desired visualizations or raising questions about what kinds of visualizations might be the most useful to target audiences.

4.1 Choosing Visualizations

During our demonstration sessions, we provided users with a variety of different visualization examples, inviting them to think about how such visualizations could be applied to collections information. Currently available web services provide a number of different kinds of visualizations, but often these have been geared towards other kinds of common statistical data – and not the rich textual descriptions commonly found in cultural heritage repositories. This proved challenging on two different levels. Firstly, many of the participants were familiar with text-based collection descriptions and had not previously considered how a visual representation of collection-level information might be used. The challenge here was not only to consider what attributes of collections would be most useful, but how those attributes would be visualized. The IMLS DCC prototype has relatively simple maps, bar charts and slightly more exotic treemap and bubble chart visualizations that were easily grasped by participants.

Additional research is needed to understand how to best provide users with multiple levels of understanding about a collection. Our discussion with participants also suggests that offering dynamic and user-centered ways to view collections. As designers we may not be able to predict how an individual user might want to see subjects across a time scale or the relationship of titles to item types [24].

Furthermore, additional research is needed to understand how these types of visualizations can be blended into traditional search and retrieval interfaces. The prototypes we have been exploring rely on somewhat static and pre-computed visualizations based on established "collection" organizations. However, participants suggested that such visualizations would also prove useful for looking across collections or to understand the contours of a search result set.

4.2 Metadata Quality (Input)

Within the cultural heritage community, issues surrounding metadata quality have been the focus of research for some time. While much of this work has argued for ways to improve search and retrieval services, metadata quality also has a significant impact on the quality and usefulness of visualizations for end users. Developing a visualization based on collection-level metadata created by IMLS DCC was relatively easy because of the metadata's consistency. However we were unable to quickly develop prototypes that worked for any arbitrary OAI-PMH set available through IMLS DCC. While individual OAI-PMH sets could be visualized, internal inconsistencies required significant human intervention to prepare the metadata for visualization. Solutions that worked for one set did not apply to adjacent sets, inhibiting the development of a generalized OAI-PMH visualization service. While the data from the Metropolitan Museum of Art represents a single institution, inconsistencies emerged across data from the nineteen different departments. Within the museum, field names, content and metadata semantics are still vigorously debated from schema development through data entry. The problems with a single institution are compounded across the multiple institutions found in IMLS DCC and maybe further exacerbated by crosswalks from local schema to common OAI-PMH Dublin Core.

A further challenge to dashboard visualizations comes from the services themselves. Because each of the APIs and web services may transport their data in any number of formats, dashboards and aggregators need to be able to translate metadata from its native format into formats required for a particular service and then further package output for use by other visualization components. In the Museum Pipes applications, JSON has been the preferred ingest language because of its nature as a self contained, lightweight interchange format. More expressive or appropriate data standards, e.g. SKOS, have not been very helpful since their implementation still needs to be examined (what fields are used, how they map to other data standards in the same tool) to assure consistent use. RSS has been a consistent choice for output from the Museum Pipes tools. Again, what it loses it expressivity or context specificity it makes up for in reusability of content across contexts – on web pages, in Java applications, or as an easily ingestible format by other services.

5. Future Research

The research that we have undertaken has been ambitious. While information visualizations and dashboards have increasingly entered the mainstream they are still novel products for cultural heritage materials. Our challenge is not only to re-imagine how aggregate information about collections might be presented but also how it might be done differently, for different audiences and different purposes. One approach to this problem could be the kind of in-depth analysis and requirements capture that precedes the development of high-fidelity prototypes (that also come with high costs in both time and funding). But lower-barrier mashup techniques allow us to more rapidly explore the design space than in the past. In essence building visualizations that help us decide what to visualize becomes a viable research strategy. The two cases presented here exploit the possibilities of tightly integrating rapid analysis and rapid design by exploiting the availability of online data and processing resources.

Several important design implications emerge from this initial work:

- The initial prototypes developed here will be useful for guiding ongoing discussion about collection dashboards. Early demonstrations that solely relied on paper prototypes left participants struggling with the concept. Later demonstration (with rudimentary mockups) were able to more quickly focus on

the possibilities and problems of visualizing collections information.

- Although it is possible to create prototypes with good metadata, inconsistent, incoherent and "unsharable" metadata creates significant challenges to creating useful visualizations.
- Even so, good metadata does not guarantee good visualizations. Adherence to a consistent format is not enough to produce a compelling visualization. While an individual repository may internally offer quality metadata, the semantics of metadata across collections/repositories prevents developing aggregation-level visualizations based on item-level metadata. This inhibits the ability of users to usefully make comparisons between and among collections.
- Constraints on metadata schema are good for visualization. The more straightforward the schema for item-level descriptions, the more likely it is to be usable across a wide range of tools. This can run counter to the goals of those writing item-level descriptions for subject matter experts, e.g. expressibility, completeness.
- Collections of different types of materials (e.g. books in the library, manuscripts in the archive or objects in the museum) may only relate along a single property and have other properties that are unique to the type of material. Additional research is needed to understand how to visualize these unique properties alongside shared properties.
- The challenges presented by visualizing heterogeneous metadata offer another possible use-case for visualizations - as a diagnostic tool for metadata creators and their supervisors. Just as financial information dashboards alert traders to fluctuations in the market, integrating a dashboard into metadata creation workflows could assist with quality control efforts. Likewise they may assist service providers who are aggregating OAI-PMH metadata and need to identify how newly harvested sets fit within an existing aggregation.

This research also raised an important distinction. Do these visualizations represent the features of a cultural heritage collection or are they better representations of metadata features? Addressing this question has implications not only for the development of collection dashboards, but also for the properties we choose to include in item-level and collection-level metadata schema and publicly shared metadata.

6. Acknowledgements

IMLS DCC portions of this research were supported by a 2007 IMLS National Leadership Research and Demonstration grant LG- 06-07-0020-07.

7. REFERENCES

[1] Duff, W.M. and Johnson, C.A. Accidentally found on purpose: Information-seeking behaviors of historians in archives. Library Quarterly 72, 4 (2002), 472-496.

[2] Tibbo, H.R. Primarily history: historians and the search for primary source materials. Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, (2002), 1–10. 1.

[3] Lee, H.L. What is a collection? Journal of the American Society for Information Science 51, 12 (2000), 1106–1113.

.

[4] Foulonneau, M., Cole, T., Habing, T.G., and Shreeves, S.L. Using collection descriptions to enhance an aggregation of harvested item-level metadata. Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries., ACM Press (2005), 32-41.

[5] Viegas, F.B., Wattenberg, M., Van Ham, F., Kriss, J., and McKeon, M. Manyeyes: a site for visualization at internet scale. IEEE Transactions on Visualization and Computer Graphics 13, 6 (2007), 1121.

[6] http://vizlab.nytimes.com

[7] Heer, J. and Boyd, D. Vizster: Visualizing online social networks. Proceedings of the 2005 IEEE Symposium on Information Visualization, (2005), 33–40.

[8] http://www.data.gov

[9] Greene, S., Marchionini, G., Plaisant, C., and Shneiderman, B. Previews and overviews in digital libraries: Designing surrogates to support visual information seeking. Journal of the American Society for Information Science 51, 4 (2000), 380–393.

[10] Derthick, M. Exploring Meta-Data Associations with Bungee View. (2007). http://www.cs.cmu.edu/~mad/2007InfoVisContest/BungeeView.pdf

[11] Kramer-Smyth, J., Nishigaki, M., Anglade, T. ArchivesZ: Visualizing Archival Collections. http://archivesz.com/ArchivesZ.pdf

[12] The Visible Archive http://archivesz.com/ArchivesZ.pdf

[13] Few, S. *Information Dashboard Design.* O'Reilly, Cambridge, MA, 2006.

[14] Legrady, George Making Visible the Invisible: Seattle Library Data Flow Visualization. Digital Culture and Heritage. Proceedings of ICHIM05.Paris. (2005).

[15] http://dashboard.imamuseum.org

[16] Stein, R. Transparency and Museums – Walking the Talk. Indianapolis Museum of Art Blog. http://www.imamuseum.org/blog/2009/11/03/transparency-and-museums/

[17] Jones, M.C., Floyd, I. R., Twidale, M. B., Patchworks of Open-Source Software: High-Fidelity Low-cost Prototypes. In St. Amant, K. and Still, B. *The Handbook of Research on Open Source Software*. ISR, Hersey, PA, (2007). 126-140.

[18] IMLS DCC Opening History http://imlsdcc.grainger.uiuc.edu/history/

[19] Palmer, C. L., Knutson, E.M., Twidale, M.B., Zavalina. O. Collection Definition in Federated Digital Resource Development In *Proc ASIST 43*. Austin, TX.

[20] Twidale, M., and Urban, R. J. Usability Analysis of the IMLS Digital Collection Registry, 2005. http://imlsdcc.grainger.uiuc.edu/3YearReport/docs/UsabilityReport1.pdf

[21] Buckland, M., et al., Access to Heritage Resources Using What, Where, When, and Who, in J. Trant and D. Bearman (eds.). *Proceedings, Museums and the Web 2007.*

[22] http://www.richardjurban.net/dashboard/

[23] http://museumpipes.wordpress.com

[24] Card, S. Mackinlay, J. and Shneiderman, B. *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufman (1999).

The Impact of Outliers: Practice Theories and Informetrics

Betsy Van der Veer Martens
University of Oklahoma
4502 E. 41st St.
Tulsa OK 71435
(918) 660-3376
bvmartens@ou.edu

ABSTRACT
This comparative case study explores the impact of four "practice theories" in the separate domains of finance, military strategy, nursing, and theology, and discusses potential "outputs" in each field that might be developed into new metrics to enrich the current practice of informetrics.

Keywords
impact, informetrics, innovation, practice, theory

1. INTRODUCTION
Despite the development of important new metrics for science and technology information [7], informetrics continues to retain certain blind spots in respect to analyzing the impact of information. One of these is the implicit assumption in much analysis that there is a single theory-driven peer-reviewed publication "point of origin" for research advances, which may be tracked by the use of appropriate citation metrics and that will also eventually translate into practical innovations that can then be further measured by appropriate patent metrics [73]. However, this fails to represent the reality of "knowledge translation" today in many fields [47].

The existing emphasis on disciplinary "citation culture" [68] and widespread reliance on the Thomson Reuters databases as a sole source of data for informetric analysis [44] has also meant that the role of different kinds of practice and practitioners in the development and diffusion of theories of all kinds has been largely ignored. The overwhelming focus on citation and patent counts disregards the fact that "practices" and "processes" may be as important as "products" in many domains, and that their "outputs" may not be captured by these measures. This has already been recognized by knowledge management scholars in particular, but most of their emphasis has been on ways that organizations can appropriate the tacit knowledge of employees rather than the analysis of impacts on a broader scale [63]. Theoretical development and diffusion outside "traditional" scholarly channels remain largely a black box to informetric investigation [36].

Crowley, one of the few in the LIS field who has recognized the importance of theorizing from practice, also associates it with the tacit knowledge of practitioners that is to be identified and codified by academic researchers in order to be properly utilized [17]. However, the theories described in the present study are clearly more than the "tacit" ones of individuals unaware of their own knowledge. This exploratory research investigates the impact of practitioner-generated theories through a comparative case study of four different theories in the fields of finance, health, military science, and theology. These four fields of practice continue to present issues of pressing importance today, and the choice of these particular theories for examination was made precisely because they did not emerge in the conventional way from disciplinary journal literature: rather, they were innovative ideas developed and diffused by practitioners in these varied fields for other, practical purposes, and only subsequently emerged in the scholarly literature through citations by academics. Their impact both on the scholarly literature and on their own domain of practice is the subject of this research. New measures of impact are also discussed in the context of each case study.

2. THE PROBLEM OF PRACTICE
A particular challenge for this project has been the vexed question of defining a "practice" theory. While those in the physical sciences may associate "practice" with Pickering's account of scientists' "mangle of practice" [52] and those in the social sciences with Schön's account of professionals' "reflective practice" [66], the so-called "practice turn" in the 1990s initiated a much wider view of the "practice field" as "embodied, materially mediated arrays of human activity centrally organized around shared practical understandings" [65]. There is great variation among these "theories of practice" originating in both sociology and philosophy, with much emphasis on the rules and skills, both explicit and implicit, that define a particular set of practices within a community, and, not unexpectedly, some criticism that the "practice turn" has been a misguided one for theorists [81].

In a cogent argument on behalf of practitioners themselves, Polkinghorne claims that the growing imposition of theory-driven "best practices" and "evidence-based practice" on professionals in many fields is robbing them of the right and obligation to exercise necessary judgment and reflection in their own practice [55]. And, finally, Knorr Cetina maintains that "The notion of a knowledge society suggests that knowledge-centered practice . . . is more dynamic, creative, and constructive than the current definition of practice as rule-based routines or embodied skills suggests" [31].

In accordance with Knorr Cetina, Pickering, Polkinghorne, and Schön, I maintain that, while practices are certainly not theories, practices can and do generate new theories in the minds of those who practice and who reflect on those practices. The four practitioner theorists studied here include a U.S. Air Force fighter pilot (Boyd), a Dominican priest (Gutiérrez), a registered nurse (Orem), and a Wall Street trader (Treynor). The work, especially the written work, of these "dynamic, creative, and constructive" practitioners is what I will term "practice theory." The impact of such "practice theories" on both "practice" and "theory" is the subject of this research.

3. METHODOLOGY

The methodological framework employed for this study is the comparative case study method [59], as that allows each of the four cases studied to contribute equally towards illuminating a general model. Though chosen from different fields of practice, the theories selected for this research originated in a single decade (the 1970s) in order to provide sufficient history for the comparative analyses, as suggested by Martens and Goodrum [41]. All of these theories are being actively utilized at present in one or more domains.

Additionally, these theories provided numerous textual artifacts such as textbooks, briefings, homilies, accounting rules, taxonomies, professional articles, and popular books to enrich this study. Specific methodologies employed included standard bibliometric analytic techniques for tracing the diffusion of theories through citation networks [19], content analytic techniques for tracing the diffusion of theories through other networks [82], and conceptual analytic techniques for studying the theoretical texts themselves [51]. In addition, at least partial bibliometric or diffusion studies of these theories have been performed by others, and the results of these studies helped to form the conclusions presented here.

4. CASE STUDIES

The specific cases examined are taken from work in finance [79], military strategy [9], nursing [49], and theology [28]. All references to the works by the four theorists in this study were identified and downloaded from the Thomson Reuters databases to form the corpus of the scholarly citations to be analyzed for each theory. The data examined included the text of the theories themselves, the text of many of the scholarly articles citing the theories selected from each decade beginning with the theory's original publication, as well as selected text from a variety of other materials that employed the theories.

4.1 Self-Care Deficit

Nurse educator Dorothea E. Orem developed the "self-care" theoretical framework for the practice of nursing over the course of the past four decades. Her framework is comprised of three subsidiary theories that involve the interlocking concepts of self-care, self-care deficit, and nursing systems. These theories deal with the various levels of self-care that people may or may not be able to provide for themselves during the lifespan, the associated needs for nursing care, and the systems of care that are thus required in that particular individual's environment.

Orem's work was grounded in her experience in nursing practice, representing her attempt to present a formal conceptualization of nursing as a domain during a period when the practice of nursing lacked much explicitly nursing-oriented theorizing to support a nursing curriculum [75]. This is exemplified by the fact that Orem's theory was first published as a nursing textbook that has now gone through six editions, the primary source for all subsequent references to her work [15].

As a practice discipline, nursing has spent a great deal of effort in considering the importance and implications of such "situation-based" theorizing [43]. However, the bibliometric study of Orem's theory is complicated by the fact that until recently the contents of only a very few peer-reviewed nursing journals were routinely input into the predecessors of Thomson Reuters's current databases. Most bibliometric analysis of nursing theories are, therefore, performed by using the CINAHL indices [70]. Almost all of the citations to Orem's theory appear in nursing or health-related journals. While nursing imports a wide variety of theories from other fields, it is far less apt to export its own [1].

In addition to the current study, three bibliometric studies of self-care deficit theory [6, 71, 76] have indicated that, while citations to Orem's work clearly indicate that it is a foundational theory in nursing, empirically-based studies that test its constructs are surprisingly few and largely superficial, given the number of references to the theory in published journal articles. This observation is clearly related to the well-known difficulties of practice theory, especially in reference to the vulnerable populations which are the central "subjects" of nursing research, with regard to the norms of clinical and scientific research practice [78].

This is an ongoing issue in nursing, exacerbated by the increasing influence of the requirements for "evidence-based" research in the medical field [40]. Although some nursing theorists are dismayed by its "colonization" of nursing [56], evidence based on the practice of nursing itself may also provide an opportunity to better employ the nursing diagnostic taxonomies constructed from Orem's theoretical framework [22]. While nursing-specific taxonomies continue to be both complex [4] and contested [8], their use may represent a necessary compromise in the face of the increased pressure for accountability at all levels in the health sector. Properly de-identified to protect individual patient privacy, the use of diagnostic and intervention data by nurses may also serve to generate more useful evidence from actual nursing practice to better inform nursing theory and, accordingly, nursing practice. Information scientists interested in classification as a form of theorizing might also find opportunities to explore the impact of nursing theories through this lens of practice by using the NANDA, NIC, and NOC taxonomies [32].

4.2 Integration Hypothesis

The so-called "integration hypothesis" was introduced in *The Financial Reality of Pension Funding Under ERISA*, a 1976 book by Treynor, Regan, and Priest, one of the first examinations of the potential impact of the newly legislated Employees Retirement Investment Security Act on corporate accounting and actuarial practices with regard to pensions. Their hypothesis was that corporations would shortly begin to consider these new legally-enforced corporate liabilities (previously considered to be unenforceable "gratuities" to retiring employees) as assets to be managed within the corporate portfolio, making particular use of the new "pension put" more or less unwittingly created by the existence of the Pension Benefit Guaranty Corporation. The book created an entirely new "integrated" framework for viewing corporate pension plans, with both positive and negative ramifications for the stakeholders involved.

Treynor may be said to represent the ultimate "reflective practitioner": he is renowned for having developed (and not published) a version of the Capital Asset Pricing Model that actually preceded Sharpe's 1990 Nobel Prize-winning effort, as well as for later having inspired Fischer Black to work on the options pricing model that helped win Merton and Scholes their Nobel Prize in economics after Black's death [45]. While displaying substantial interest in portfolio theory, Treynor

consistently emphasized the practical side of finance, as he is also noted as a pioneering proponent of financial engineering departments within banks and brokerages: the so-called "quants" that have transformed American finance over the past decades [35].

A brief condensation by Treynor of the book's main points also appeared in a special issue of the *Journal of Finance* in 1977. The publication counts, therefore, include both citations to the book and to this article by Treynor. Initial scholarly citations to the book were immediate: the first was in *Harvard Business Review* [77], and continue into the present, long after the book itself has gone out of print [27, 61]. Because of its revolutionary, options-oriented view of the PBGC's unforeseen role as the captive buyer of a corporation's unfunded pension liability, the book became an almost obligatory reference in any work dealing with pension funding. Amsbachteer commented that fully understanding the implications of this practical approach to pension management still forms part of the "next frontier of portfolio theory" [2].

Whitley has chronicled the development of portfolio theory and financial economics as "a particularly interesting example of new 'occupation-oriented' scientific fields since it combines a high degree of practitioner interest and support with a high degree of theoretical abstraction and coherence, which is unusual in such fields, particularly those concerned with social phenomena" [83]. More recently, MacKenzie [37, 38] has speculated that financial engineering actually creates markets through these abstract models and the various mathematical products that embody them. Relatedly, "enterprise risk management" has become a critical part of the firm's thinking, due both to the risks involved in financial markets and the financialization of risk management itself [57].

Again, information science could provide a beneficial set of informetric tools with which to examine not only the texts, but the so-called "market devices" of algorithms, pricing models, trading protocols, financial instruments, and the aggregate data that "make" financial markets beyond the insights provided by purely econometric models [48]. Surprisingly little informetric interest has been shown so far in what is clearly a "bull market" of potential data points for the broader impact of particular financial theories.

4.3 Liberation Theology

Gutiérrez's book, *A Theology of Liberation*, was originally published in Spanish in 1971, with an English translation published by Orbis Books in 1973. Considered a founding work in so-called "liberation theology," this theory emerged from a controversial series of meetings held by Catholic bishops in Latin America regarding the role of the post Vatican-II Catholic Church in the often tortuous economic and political "development" of their countries. Gutiérrez, an attending priest, articulated his insight that theology is "critical reflection on praxis in the light of the word of God" and that the mission of the Church regarding poor and oppressed people worldwide should be "liberatory" rather than simply "developmental." Simply stated, his theory is that such liberation should be on three levels: individual (the liberation by Christ from sin), social (the liberation of entire communities from the selfish refusal to love one's neighbors) and global (the liberation of humankind from the historical acceptance of misery, despoilation, and alienation as a "natural" condition of human existence as propounded by the Church). Gutiérrez has been continually engaged in dialogue regarding his theological writings with the Vatican's Congregation for the Doctrine of the Faith, most particularly with its head (now Pope Benedict XVI), who published an "Instruction" regarding the relationship of so-called "liberation theology" to accepted Church doctrine [60].

The two English editions of the book reached an audience far beyond the pastoral community in Peru. "Liberation theology" was perceived to carry strong political implications, due to its connections with Marxist thought, especially by the use of the term "praxis." Although Gutiérrez's use of the term was intended to denote "Christian praxis" as "orthopraxis" (right practice, consistent with Catholic orthodoxy), many supporters of "liberation theology" were indeed actively Marxist in their political aspirations [69]. The decade after the birth of "liberation theology" saw the increasing involvement of Catholic clergy in the struggle against political repression in South America [5]. However, as the political situation changed, many "liberatory" bishops were replaced by less radical ones, and many liberation theologians were rebuked for their writings by the Vatican [34].

Citations to Gutiérrez's book, however, appear to continue strongly, even though a more comprehensive bibliometric approach would entail using the ATLA (American Theological Library Association) religion database, as similar concerns about the extent of the Thomson Reuters indexing of peer-reviewed theological journals exist as the ones described above for nursing journals. While the majority of the citations are in theological journals, particularly Catholic-oriented ones, the theory has also been used in a wide variety of contexts: for example, behavioral science [29], law [58] political science [26], and social psychology [13].

While Gutiérrez's work has received sufficient scholarly citations to qualify it as a citation classic, its real *raison d'etre* was to challenge the magisterium (the official teachings of the Catholic Church) regarding the Church's special obligation to the poor, which is why it has received so much attention from the Vatican over the decades. While the so-called "ordinary universal magisterium" involves revealed matters of faith and is properly promulgated only by the Pope and the college of bishops, the "ordinary magisterium" is the one through which theologians raise contemporary issues of innovation and interpretation that may affect both the theory and practice of the faith [74].

What bibliometric analysis has not done is to examine the intricate interplay of how these intellectual innovations by theologians can influence and be influenced by the magisterium [23]. The critical role of theological literature in this process of "complementary charisms" [64] is largely unexplored by secular scholars. Further, the ongoing importance of religious teachings has been neglected, but clearly their impact has been felt both politically and socially, not only through Catholicism and in Latin America, but through Islam and in the Middle East. Gill [25] suggests that the study of theories such as liberation theology and their diffusion both inside and outside doctrinary channels would also allow the building of more general theories of how such ideas and institutions interact.

4.4 OODA Loop

Boyd's "OODA loop" theory represents an intriguing case of a highly specialized practitioner theory that went largely unnoticed

in the scholarly literature for the first two decades of its existence, while becoming increasingly influential in the military domain in which it was practiced, and which has now given rise to new uses in several disciplines. This "observation, orientation, decision, action" systems-based approach to strategic cognition as a time-based theory of conflict was grounded in Boyd's practical knowledge as a Navy fighter pilot of how to outmaneuver an enemy in aerial combat [50].

Boyd helped to codify his tacit knowledge of air maneuvers in a series of Air Force reports, but the OODA loop theory was an separate outgrowth of his realization that this tacit knowledge, synthesized with existing explicit knowledge about strategy, could also be applied to operational service practices [30]. This so-called "operational art" occupies the middle ground between tactics and strategy, all of which are considered uniquely military types of theorizing [21]. The primary mechanism for dissemination of the OODA loop was Boyd's famous in-person six-hour "briefings" to various levels of military command, illustrated by his series of slides, which intensified after his formal retirement from the military [46].

Boyd's efforts in promoting the theory throughout the Pentagon and various branches of the services eventually resulted in the OODA loop achieving what Latour [33] would term "black box" status: that is, the OODA loop itself has become reified and is often referred to without mentioning Boyd at all. Some of Boyd's associates eventually began to employ his ideas in a business context, and the OODA loop is now frequently referred to in popular management literature as well [62].

Scholarly citations to the OODA loop via Boyd's unpublished papers began in 2000, with such varied applications as software agents, fleet navigation, trauma medicine, and disaster management. One of the manifest difficulties in citing Boyd's work is that he continued to work on the theory throughout his lifetime without producing a definitive published document, and so there is an unusual variation in references to his works, ranging from the "Destruction and Creation" document [9] to various other permutations of his presentations and reports about the presentations. Clearly, however, the most critical views of the OODA loop come from within the military itself, which can be considered the "peer review" process for this theory: other scholarly uses appear to incorporate the OODA loop without similar reservations.

The OODA loop diffused through a variety of military channels [3] and is recognized as a foundational "timing" concept in operational art [18]. Although it has been suggested that the OODA loop theory is of diminished utility in so-called "fifth-generation warfare" [67], other commentators argue that this opinion is grounded in a superficial understanding of the theory [54]. Despite its manifest flaws as a cognitive model [12], the OODA loop remains prominent in so-called "Command and Control" military doctrine [10]. The OODA loop remains one of the few contemporary examples of "bottom-up" theoretical innovations in military doctrine, which are the most understudied examples of military innovation, as most diffusion studies in this field tend to focus on examples of historical interest, high technology, or grand strategy [20].

Although there has been much interest in both research and development inputs to the military [11] and research and development outputs by the military [53], there has been surprisingly little interest in how ideas diffuse through the military itself, with the exception of a small cadre of theorists in strategic studies [16]. This may be due to the perceived difficulty in gaining access to the necessary documents, either because they are classified [24], in the under-indexed "gray literature" [14], conform to different stylistic conventions than those in academic writing [42], or simply because it is not generally understood that, as "weapons that think" [39] military organizations engage in practical theorizing at all levels of leadership [72]. These are issues related to innovation and information that will be increasingly pressing in the current era of so-called cyberwar, and information scientists could begin to examine them through the texts of military doctrines, military operational concepts, and actual military orders as these become available for research use.

5. CONCLUSIONS

Most informetric studies begin rather than end with the actual citation counts of the theories analyzed. I have deliberately placed Figure 1 at the end of this article in the hopes that the discussion here has illuminated some of the weaknesses of citation analysis in considering the overall importance and impact of practice theories, especially when utilizing a single data source such as Thomson Reuters. The number of citations listed for each of these theories fails to indicate their impact on their own fields of practice, nor, presumably, on society at large. These impacts vary by case:

- Orem's work has had significant impact on both theory and practice in the emerging field of modern nursing, but little influence on the broader medical field
- Treynor's work has had roughly equivalent influence on both pension portfolio theory and pension fund practice
- Gutiérrez's work has had more influence both on theological theory in general and on the theories of other disciplines than on the practices of the Catholic Church
- Boyd's work has had significant impact on the practices of his own military hierarchy and those of related organizations, while having some very limited influence outside the military

Clearly, the impact of practitioner-generated theories can be as disparate as those of academically-generated theories, ranging from non-existent to highly influential. The current lack of interest and exploration in these areas by informetricians is somewhat surprising, given the increasing importance of innovations in these particular fields.

Tukey [80] famously urged researchers to explore what statistical outliers might reveal about the data being analyzed. Practice theories continue to be the "outliers" in the assessment of innovation and impact. Informetricians might well take into account the methods of dissemination and the contributions of practice theories as they contemplate new indicators and indicator theories [84].

FIGURE 1. DATA ANALYSIS OF PRACTICE THEORIES

Theory:	Citation counts in Thomson Reuters:	Data analyzed:	Primarily diffused via:	Possible "practice" outputs:
Self-care deficit [Nursing] Orem	781	Nursing taxonomies Nursing textbooks Professional articles Scholarly articles	Nursing textbooks	Nursing diagnoses Nursing interventions
Liberation theology [Theology] Gutiérrez	456	Popular articles Popular books Religious texts Scholarly articles Theological books	Religious literature (including homilies)	Religious doctrines
Integration hypothesis [Finance] Treynor	96	Financial textbooks Professional articles Scholarly articles	Financial publications	Portfolio holdings
OODA loop [Military] Boyd	9	Management books Military briefings Professional articles Scholarly articles	Military briefings Management books	Military doctrines Business strategies

6. REFERENCES

[1] Allen, P., Jacobs, S. K., and Levy, J. R. 2006. Mapping the literature of nursing: 1996-2000. Journal of the Medical Library Association, 94, 2, 206-220.

[2] Ambachtsheer, K. 2005. Beyond portfolio theory: The next frontier. Financial Analysts Journal, 61, 1, 29-33.

[3] Angerman, W. S. 2004. Coming full circle with Boyd's OODA loop ideas: An analysis of innovation diffusion and evolution. Master's thesis. Air Force Institute of Technology, School of Engineering and Management,Wright-Patterson Air Force Base.

[4] Beckstead, J. W. 2009. Taxonomies of nursing diagnoses: A psychologist's view. International Journal of Nursing Studies, 46, 295-301.

[5] Berryman, P. 1987. Liberation theology: Essential facts about the revolutionary movement in Latin American and beyond. New York: Pantheon Books.

[6] Biggs, A. 2008. Orem's self-care deficit nursing theory: Update on the state of the art and science. Nursing Science Quarterly, 21, 200-206.

[7] Bollen, J., Van de Sompel, H., Hagberg, A., and Chute, R. 2009. A principal component analysis of 39 scientific impact measures. PloS One. Available: http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0006022

[8] Bowker, G. C., Star, S. and Spasser, M. 2001. Classifying nursing work. Online Journal of Issues in Nursing, 6, 2. Available: http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume62001/No2May01/ArticlePreviousTopic/ClassifyingNursingWork.aspx

[9] Boyd, J. R. 1976. Destruction and creation. Unpublished presentation delivered 23 Mar 1976, Air Forces University.

[10] Brehmer, B. 2005. The dynamic OODA loop: Amalgamating Boyd's OODA loop and the cybernetic approach to command and control. Presented at the 10th International Command and Control Research and Technology Symposium. June 13-16, 2005, McLean, Virginia.

[11] Brooks, S. G. 2005. Producing security: Multinational corporations, globalization, and the changing calculus of conflict. Princeton: Princeton University Press.

[12] Bryant, D. J. 2006. Rethinking OODA: Toward a modern cognitive framework of command decision making. Military Psychology, 18, 3, 183-206.

[13] Burton, M. and Kagan, C. 2005. Liberation social psychology: Learning from Latin America. Journal of Community and Applied Social Psychology, 15, 63-78.

[14] Chapman, B. 2009. Military doctrine: A reference handbook. Santa Barbara CA: Praeger Security International.

[15] Clarke, P. N., Allison, S. E., Berbigilia, V. A., and Taylor, S. G. 2009. The impact of Dorothea E. Orem's life and work. Nursing Science Quarterly, 23, 41-46.

[16] Cohen, E. 2009. Change and transformation in military affairs. In B. Loo, Ed., Military transformation and strategy: Revolutions in military affairs and small states. pp. 15-26. London: Routledge.

[17] Crowley, B. 1999. Building useful theory: Tacit knowledge, practitioner reports, and the cult of LIS inquiry. Journal of Education for Library and Information Science, 40, 282-295.

[18] Cunningham, K. and Tomes, R. R. 2004. Space-time orientations and contemporary political-military thought. Armed Forces and Society, 31, 119-140.

[19] De Bellis, N. 2009. Bibliometrics and citation analysis: From the Science Citation Index to cybermetrics. Lanham MD: Scarecrow Press.

[20] Eliason, L. C. and Goldman, E. O. 2003. Theoretical and comparative perspectives on innovation and diffusion. In E. O. Goldman and L. C. Eliason, Eds., The diffusion of military technology and ideas. pp. 1-30. Stanford CA: Stanford University Press.

[21] English, J. 1996. The operational art: Developments in the theories of war. In B. J. C. McKercher and M. A. Hennessy, Eds., The operational art: Developments in the theories of war. pp. 7-27. Westport CT: Praeger Publishers.

[22] Fawcett, J. 2003. Orem's self-care deficit nursing theory: Actual and potential sources for evidence-based practice. Self-Care, Dependent-Care, and Nursing, 11, 11-16.

[23] Gaillardetz, R. R. 1997. Teaching with authority: A theology of the Magisterium in the Church. Collegeville MN: The Liturgical Press.

[24] Galison, P. 2004. Removing knowledge. Critical Inquiry, 31, 229-243.

[25] Gill, A. 2002. The study of liberation theology: What next? Journal for the Scientific Study of Religion, 41, 87-89.

[26] Gill, A. J. 1994. Rendering unto Caesar: Religious competition and Catholic political strategy in Latin America, 1962-79. American Journal of Political Science, 38, 403-425.

[27] Glaum, M. 2009. Pension accounting and research: A review. Accounting and Business Research, 39, 273-311.

[28] Gutiérrez, G. 1973. A theology of liberation. Maryknoll, NY: Orbis Books.

[29] Hagan, J. 2006. Making theological sense of the migration journey from Latin America: Catholic, Protestant, and interfaith perspectives. American Behavioral Scientist, 49, 1554-1573.

[30] Hammond, G. T. 2001. The mind of war: John Boyd and American security. Washington: Smithsonian Books.

[31] Knorr Cetina, K. 2001. Objectual practice. In T. R. Schatzki, K. Knorr Cetina, and E. von Savigny, Eds., The practice turn in contemporary theory. pp. 175-188. London: Routledge.

[32] Kumar, C. P. 2007. Application of Orem's self-care deficit theory and standardized nursing languages in a case study of a woman with diabetes. International Journal of Nursing Terminologies and Classifications, 18, 3, 103-110.

[33] Latour, B. 1987. Science in action: How to follow scientists and engineers through society. Cambridge MA: Harvard University Press.

[34] Lernoux, P. 1989. People of God: The struggle for world Catholicism. New York: Viking Books.

[35] Lindsay, R. R. and Schachter, B. 2007. How I became a quant: Insights from 25 of Wall Street's elite. Hoboken NJ: John Wiley.

[36] Luukkonen, T. 1997. Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. Scientometrics, 38, 1, 27-37.

[37] MacKenzie, D. 2006. An engine, not a camera: How financial models shape markets. Cambridge MA: MIT Press.

[38] MacKenzie, D. 2007. Is economics performative? Option theory and the construction of derivatives markets. In D. MacKenzie, F. Muniesa, and L. Siu, Eds., Do economists make markets? On the performativity of economics. pp. 54-86. Princeton: Princeton University Press.

[39] Mandeles, M. D. 2005. The future of war: Organizations as weapons. Washington: Potomac Books.

[40] Mantzoukas, S. 2009. The research evidence published in high impact nursing journals between 2000 and 2006: A quantitative content analysis. International Journal of Nursing Studies, 46, 479-489.

[41] Martens, B. V. and Goodrum, A. 2006. The diffusion of theories: A functional approach. Journal of the American Society for Information Science and Technology, 57, 330-341.

[42] McIntosh, W. A. 2003. Guide to effective military writing. 3rd ed. Mechanicsburg PA: Stackpole Books.

[43] McKenna, H. P. 1997. Theory and research: A linkage to benefit practice. International Journal of Nursing Studies 34, 6, 431-437.

[44] McRoberts, M. H. and MacRoberts, B. R. 2010. Problems of citation analysis: A study of uncited and seldom-cited influences. Journal of the American Society for Information Science and Technology, 61, 1, 1-13.

[45] Mehrling, P. 2005. Fischer Black and the revolutionary idea of finance. Hoboken NJ: John Wiley and Sons.

[46] Meilinger, P. S. 2000. The historiography of airpower: Theory and doctrine. The Journal of Military History, 64, 467-501.

[47] Meyer, M. 2009. Measuring knowledge translation in the S&T environment. Presentation at the Association for Learned & Professional Societies Publishing seminar, London, 15 June 2009. Available: http://www.alpsp.org/ngen_public/article.asp?id=0&did=0&aid=44958&st=impact metrics

[48] Muniesa, F., Millo, Y., and Callon, M. 2007. An introduction to market devices. In M. Callon, Y. Millo, and F. Muniesa, Eds., Market devices. pp. 1-12. Oxford: Blackwell Publishing.

[49] Orem, D. E. 1971. Nursing: Concepts of practice. St. Louis: Mosby.

[50] Osinga, F. P. B. 2007. Science, strategy and war: The strategic theory of John Boyd. New York: Routledge.

[51] Palmquist, M. E., Carley, K. M., and Dale, T. A. 1997. Two applications of automated text analysis: Analyzing literary and non-literary texts. In C. Roberts, Ed., Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts. pp. 171-189. Hillsdale NJ: Lawrence Erlbaum Associates.

[52] Pickering, A. 1995. The mangle of practice: Time, agency and science. Chicago: University of Chicago Press.

[53] Pierce, T. C. 2004. Warfighting and disruptive technologies: Disguising innovation. New York: Frank Cass.

[54] Polk, R. B. 2000. A critique of the Boyd theory: Is it relevant to the Army? Defense Analysis, 16, 3, 257-276.

[55] Polkinghorne, D. E. 2004. Practice and the human sciences: The case for a judgment-based practice of care. Albany: State University of New York Press.

[56] Porter, S. and O'Halloran, P. 2009. The postmodernist war on evidence-based practice. International Journal of Nursing Studies, 46, 740-478.

[57] Power, M. 2005. Enterprise risk management and the organization of uncertainty in financial institutions. In K. Knorr Cetina and A. Preda, Eds., The sociology of financial markets. pp. 250-268. New York: Oxford University Press.

[58] Quinn, K. P. 2000. Viewing health care as a common good: Looking beyond political liberalism. Southern California Law Review, 73, 2 , 277-375.

[59] Ragin, C. C. 1987. The comparative method: Moving beyond qualitative and quantitative methods. Berkeley: University of California Press.

[60] Ratzinger, J. C. 1984. Instruction on certain aspects of the "theology of liberation." Congregation for the Doctrine of the Faith. Available: http://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_con_cfaith_doc_19840806_theology-liberation_en.html

[61] Rauh, J. D. 2009. Risk shifting versus risk management: Investment policy in corporate pension plans. Review of Financial Studies, 22, 2697-2733.

[62] Richards, C. 2004. Certain to win: The strategy of John Boyd applied to business. Bloomington: Xlibris.

[63] Roberts, J. 2001. The drive to codify: Implications for the knowledge-based economy. Prometheus, 19, 2, 99-116.

[64] Salzman, T. A. and Lawler, M. G. 2009. Theologians and the magisterium: A proposal for a complementarity of charisms through dialogue. Horizons, 36, 1, 7-31.

[65] Schatzki, T. R. 2001. Introduction: Practice theory. In T. R. Schatzki, K. Knorr Cetina, and E. von Savigny, Eds., The practice turn in contemporary theory. pp. 1-14. London: Routledge.

[66] Schön, D. A. 1983. The reflective practitioner: How professionals think in action. New York: Basic Books.

[67] Scott, W. J., McCone, D. R., and Mastroianni, G. R. 2009. The deployment experiences of Ft. Carson's soldiers in Iraq: Thinking about and training for full-spectrum warfare. Armed Forces and Society, 35, 460-476.

[68] Skilton, P. F. 2006. A comparative study of communal practice: Assessing the effects of taken-for-granted-ness on citation practice in scientific communities. Scientometrics, 68, 73-96.

[69] Smith, C. 2001. The emergence of liberation theology: Radical religion and social movement theory. Chicago: University of Chicago Press.

[70] Smith, D. R. and Hazelton, M. 2008. Bibliometrics, citation indexing, and the journals of nursing. Nursing and Health Sciences, 10, 260-265.

[71] Spearman, S. A., Duldt, B. W., and Brown, S. 2003. Research testing theory: A selective review of Orem's self-care theory, 1986-1991. Journal of Advanced Nursing, 18, 1626-1631.

[72] Sternberg, R. et al. 2000. Practical intelligence: An example from the military workplace. In R. Sternberg, Ed. , Practical intelligence in everyday life. pp. 162-206. New York: Cambridge University Press.

[73] Sternitzke, C. 2009. Patents and publications as sources of novel and inventive knowledge. Scientometrics, 79, 551-561.

[74] Sullivan, F. A. 1996. Creative fidelity: Weighing and interpreting documents of the Magisterium. Mahwah NJ: Paulist Press.

[75] Taylor, S. G. 2007. The development of self-care deficit nursing theory: An historical analysis. Self-Care, Dependent-Care, and Nursing, 15,1, 22-25.

[76] Taylor, S. G., Geden, E., Isaramalai, S., and Wongvatunyu, S. 2000. Orem's self-care deficit nursing theory: Its philosophic foundation and the state of the science. Nursing Science Quarterly, 13, 104-110.

[77] Tepper, I. 1977. Risk vs. return in pension fund investment. Harvard Business Review, 55, 2, 100-107.

[78] Tolley, K. A. 1995. Theory from practice for practice: Is this a reality? Journal of Advanced Nursing, 21, 184-190.

[79] Treynor, J. L., Regan, D. J., and Priest, W. W. 1976. The financial reality of pension funding under ERISA. Homewood IL: Dow Jones-Irwin.

[80] Tukey, J. W. 1977. Exploratory data analysis. Reading, MA: Addison-Wesley.

[81] Turner, S. 1994. The social theory of practices: Tradition, tacit knowledge, and presuppositions. Chicago: University of Chicago Press.

[82] White, M. D., and Marsh, E. E. 2006. Content analysis: A flexible methodology. Library Trends, 55 1 , 22-45.

[83] Whitley, R. 1986. The structure and context of economics as a scientific field. In W. J. Samuels, Ed., Research in the history of economic thought and methodology. Volume 4. pp. 179-209. Greenwich CT: JAI Press.

[84] Wouters, P. 1999. Beyond the holy grail: From citation theory to indicator theories. Scientometrics, 44,3, 561-580.

Incentives in the Wild: Leveraging Virtual Currency to Sustain Online Community

Yang Wang
Department of Informatics
University of California, Irvine
Bren Hall 5091
Irvine, CA 92697 USA
yangwang@uci.edu

Scott D. Mainwaring
People and Practices Research Group
Intel Labs
20270 NW Amberglen Ct., MS AG1-110
Beaverton, OR 97006 USA
scott.mainwaring@intel.com

ABSTRACT
The importance of incentive mechanisms has long been recognized in sustaining online communities. However, many existing community systems have rigid system-imposed incentive mechanisms that preclude user appropriations of the incentives. We draw upon recent investigations of the role of virtual currency (VC), a form of incentive, in a vibrant online community in order to highlight emergent practices that were enabled, facilitated and manifested through exchanges of virtual currency among users. Our study shows evidence that these user appropriations of VC help sustain the community and fulfill users' broader needs on these systems.

Author Keywords
Incentive, virtual currency, online community, Chinese.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Virtual economy is more than a fad for the Chinese. According to a recent report by the China Internet Network Information Center (CNNIC) [3], a 76.5% of the 55.5 million Chinese online gamers spent money on virtual items and game accounts, and the market of virtual items is valued about 10-13 billion CNY (about 1.5 billion USD). A central piece of this virtual economy jigsaw is virtual currency. Virtual currency schemes are sprouting like bamboo shoots after a spring rain in the Chinese Internet space. Online services from web portals (e.g., sina.com) to search engines (e.g., baidu.com), from sports news (e.g., titan24.com) to online games (e.g., 9you.com), from online forums (e.g., mitbbs.com) to online shopping sites (e.g., zuoz100.com), all have their own virtual currencies. Although perhaps most developed at present in Chinese contexts, large-scale virtual money systems may soon feature in Western enterprises as well (with few exceptions in online games such as World of Warcraft). For example, the largest US social networking service Facebook is testing a virtual currency system [1]. Potentially, many millions of people worldwide could use this system. We assert that virtual monetary transactions function in these online communities not only as a convenient means of micropayment (for their fee-based services), but also as an effective mechanism to engage their users and to foster the online communities. We will present early evidence from a popular Chinese online forum, MITBBS.

There are several threads of related work. One thread of work studies incentive mechanisms for *encouraging participation*. Farzan et al. [5] tested a static point-based incentive mechanism in IBM's internal social networking service Beehive and found a short-lived effect of the incentives. Cheng and Vassileva [2] demonstrated positive effects of an incentive mechanism that adapts the rewards to both users' reputation and the current needs of the community in an online learning system. Another thread of work examines the *answer quality* in community Q&A sites and their underlying incentive mechanisms. Harper et al. [6] found that fee-based Google Answers service typically provides higher-quality answers than its free counterparts (e.g., Yahoo! Answers). The third thread of work applies market mechanisms to help solve *communication problems*, e.g., spam. Hsieh et al. [8] showed that market mechanisms can be used to improve synchronous communication (e.g., instant messaging). However, market mechanisms are not necessarily beneficial to online communities. More recently, in the context of a community Q&A system, Hsieh and Counts [7] showed that while market mechanisms can improve content quality by screening out less serious questions and answers, they may also reduce social interactions and thus users' sense of community.

In our work, we investigated the culture, policies and practices of MITBBS *bi* ("coin") or MB, a virtual currency and form of incentive, in MITBBS. Unlike many system-imposed incentive mechanisms that preclude user appropriations of the incentives, MITBBS supports direct exchanges of MBs among individual users, creating an interesting gift economy and enabling a wide range of

"user-generated" rather than system-imposed incentive practices that are valuable in sustaining the online community.

MITBBS

MITBBS (a.k.a., Unknown Space) is a Bulletin Board System (BBS) started as bbs.mit.edu in 1996 by Liu Jia, who came from Mainland China to study at MIT. The site has become one the most popular sites for Chinese diasporas in North America and recognized to help them keep connections with homeland, unite themselves and navigate life in the western society, and "forge and strengthen their fragmented and hybridized cultural identity" [10]. The posts on the site are predominately in simplified Chinese, although some are in traditional Chinese and English. It moved from the mit.edu domain to its own domain mitbbs.com in 2002. The site now has over 100,000 registered users and over 200 topic boards. Topics range from China News to job hunting, from investment to medicine, from soccer to California. Although now the site is mostly used from its web-based interface via HTTP, users can still visit the system via TELNET. All posts on the site are publicly visible but only registered users can write a new post and/or reply to an existing post.

In terms of its admin structure, MITBBS has site admins who make site-wide polices and board admins who make board-specific rules for their respective topic boards. These board admins are usually self-nominated and then selected by the site admins. The site admins are people who run the business of the site, while board admins are just regular users albeit with certain privileges on their boards.

MB

MITBBS has two virtual currencies – the old MB and the new MB. MB stands for MITBBS *bi* ("coin"). It is also called *wei bi* ("fake coin") or *bao zi* ("bun"), somewhat analogously to the English colloquialism "dough" for cash. There is no official channel to exchange MBs to real currencies (e.g., US dollars) or vice versa. MBs can only be earned in the system (either as system incentives or gifts from other users). The old MB can be used for gifting and betting (e.g., on sports games) on the site. When the site unveiled its MShow service (i.e., dressing users' avatars) in late 2007, the new MB was introduced with the intent to replace the old MB. In addition to the old usages, the new MB can be used to buy virtual item for decorating users' avatars. While MBs are associated with every registered user's account, anyone on the web can see the amount of MBs that a user has. However, traces of MB transfer are not public. We chose to mainly study the new MB since it has more utilities and thus was more interesting. For the rest of the paper, we will focus on the new MB unless otherwise specified.

Our study sought to understand (1) the practices around MBs, particularly the transfer of MBs among users, and (2) how these practices impact the online community.

METHODOLOGY

We have been reading daily on MITBBS since the advent of its new MB for about 2 years. In this exploratory study, we use a content analysis approach. The texts of the analysis came from posts that explicitly mentioned MB. More specifically, we searched for 伪币 (wei bi or "fake coins") and 包子 (bao zi or "bun") using the site's search function. This yielded a total of 3155 posts by June 1st, 2009. They were summarized and categorized post-hoc into a list of key themes. This can be viewed as an ethnography of texts [4]. We also plan to conduct interviews with site users and admins.

EARNING MB

There are a number of ways to earn MBs: (1) When a user writes a new post on a board, the system gives the user and the board 0.1 MB, respectively. There are upper limits in terms of the amount of MBs that can be earned per day this way: 10 MBs for a user ID and 100 MBs for a board. (2) If the contributed post is of high quality and thus marked by the board admin, the user/author will get 10 MBs per marked post. In this context, the practice of board admin marking a post and giving its author 10 MBs is called *fa bao zi.* ("to offer a bun"). By extension, *fa bao zi* has come to mean offering MBs in general. (3) Board admins are compensated for their work in MBs. (4) Winners of the weekly fashion competitions in MShow get 50 MBs (big award) or 20 MBs (nomination award). (5) When a user contributes business information to the site's Yellow Pages, the system awards the user 10 MBs when the information is authorized. Unlike board postings, there is no daily limit of earning MBs this way. (The site also claims that this is the fastest way to earn MBs). (6) Finally, users who participate in events hosted by individual boards are awarded MBs. The exact incentive rules are determined by the particular board admins. For example, the fashion board encourages users to post photos of themselves by offering 10MBs per post. This practice of posting one's own photos that may potentially reveal his or her true identity is called *luo ben*, literally, "nude run".

In our analysis, we also discovered that MB-related policies (such as when to offer MBs, and the amount of MBs) were determined via a fairly open process. Site admins drafted site-wide MB rules and polled users for their feedbacks, and then integrated these feedbacks into the working policies. The development of board-specific MB policies follows a similar process. Indeed, many changes incorporated in the new MB were proposed by ordinary users.

USER-TO-USER MB TRANSFERS

Before the introduction of the new MB, users who wanted to transfer MBs had to use board admins as proxies. Users would transfer MBs to a topic board, and then the board admins would transfer the MBs to designated recipient(s). The new system not only allows users to directly transfer MBs among themselves, but also prohibits board admins from transferring MBs to individual users. In the new MB

scheme, board admins can only give MBs to users indirectly by marking posts on the board.

WHY USERS TRANSFER MB

Material Needs

Users often use MBs to encourage replies to their requests for resources and services. To do that, they usually mention MB in the titles of their posts. Examples range from asking questions about visa applications to soliciting shopping coupons, from asking people to sing a certain song on the music board to finding out someone's true identity (*ren rou sou shuo,* "flesh search"). These requests usually serve practical and functional needs of the requesters.

Emotional Displays

It is also a common practice for users to transfer MBs to convey their emotions. Examples include showing sympathy (e.g., to fellow snowboarding fans who got stuck in airports for days because of the bad weather), appreciation (e.g., to people who provided an electronic Chinese manual for a digital camera), apology (e.g., for offending people on the board), and happiness (e.g., for celebrating Chinese New Year).

Ren Pin (RP)

People often ask the community for blessings – when their parents are about to have visa interviews, or they have job interviews or green card applications. They may use MBs to attract and thank the repliers. This practice also relates to an interesting notion of *Ren Pin* (RP, "moral quality"). It is generally believed that if someone has good RP, she will have good luck. Posts that provide useful information or give away MBs are sometimes referred as "RP posts" – because they could help others and would in turn improve the authors' RP and bring them good luck.

Attracting Attention

It is not surprising that among the 200 or so topic boards on the site, some boards have high traffic (frequent new posts and replies) while others are unpopular. As a user ID has an MB account, a topic board also has a MB account which is managed by the board admins. Any registered user can transfer MBs to a board account, but only the board admins can indirectly transfer MBs out of the board account by marking posts. Sometimes board admins publish and mark their own *bao zi posts* (posts that give away MBs) just to attract more visitors to their board or to acknowledge sudden surge of visits on their boards.

MB for Gaming

Since the MShow service has system support for second-handed virtual item market, users can freely sell and buy used virtual items for their avatars using the MBs. We also witnessed that MBs were used as incentives in other games on the site such as the Killer Game (a.k.a., the Werewolf game). One post also showed the authors used MBs in playing Mahjong, a popular game originated in China:

> "We invited friends to play Mahjong at home during the weekend…we felt that only playing chips is not exciting enough, but we were not sure if playing cash is legal or not. A friend suggested using MITBBS MBs…After several hours of playing, everyone is still stimulated. [Using MBs] makes the game much more exciting."

Borrowing MB

Since only registered users can post on the site, when site readers (without accounts) want to ask questions on the site they can use their friends' accounts. If their questions receive little attention (few reply), they may borrow MBs from their friends to draw the community's attention

TENSIONS AND CONTROVERSIES OF MB PRACTICES

A number of tensions and controversies have emerged from the practices involving MBs.

Bao Zi Posts on the Top 10 List

Bao zi posts are posts that give away MBs and that usually mention MB in their titles. On the front page of MITBBS, there is a list of top 10 posts (based on the number of replies to the original posts). Many users complained that top 10 posts were no longer worth reading because many of them got replies because they were bao zi posts. Most of the replies are as simple as "re", serving as a placeholder for the authors to identify MB recipients. These replies are clearly not informative at all, and can be seen as a waste of time for the site readers.

Gaming MB

Since MBs are useful on the site, users like to have them. Some users tried to game the system to gain MBs. One infamous practice is that the same user uses different identities (*ma jia,* "disguise") to get MBs. The MB policies are universally in favor of distributing MBs fairly and to as many users as possible. Sometimes, a bao zi post explicitly informs its readers the amount of MBs are available for give-away and/or implies the number of users who can receive the MBs, e.g., only the first 5 repliers can get the MBs. There are various strategies that users can apply to increase their chances of getting MBs. In fact, some users even sell these strategies marketed as "MB secrets" to other users. One strategy is to use telnet rather than web-based (http) to connect to the site because users can reply posts quicker in telnet connections, thus increasing the odds of getting MBs.

Real Money Trade (RMT)

Many virtual currencies especially those associated with games have the issue of RMT. WoW gold, for example, is not officially for sale but has been wildly traded on "black" market for real money or point cards which have direct monetary value. In contrast, we observed very few cases of RMT of MB and believe an important reason is that

MITBBS is free and thus trading MB is not very profitable. When users trade MBs, they do it not for making profits:

A: Great news, I offer real service for exchanging MBs to US Dollars, current rate: USD $1.0 == 500 MB. If you are interested, transfer at least 500 MB to my MITBBS account and provide your PayPal account.

B: Why do you need MB?

C: What do you want to do?

A: I owe people Bao Zi.

A: The chemistry board is waiting for me to Fa Bao Zi. I don't have enough MB at the moment.

DISCUSSION AND CONCLUSION

Our study reveals that MB affects MITBBS at multiple levels. At the individual level, we found evidence of MB encouraging participation and increasing the likelihood of getting high-quality answers. This to some extent corroborates the existing literature of fee-based Q&A sites providing higher-quality answers than free Q&A sites [6]. MB is also instrumental in supporting users' emotional needs (e.g., seek blessings before interviews). At the board level, MB provides an incentive for ordinary users to work as board admins (MB as salary). MB is also useful in drawing more users and attention into boards, e.g., board admins give away MB or organize activities using MB as rewards. At the site level, MB makes the site more fun to play with (e.g., dressing and rating avatars). It also helps organize cross-board activities, integrate and unite different boards and sub-communities. However, MB is not always what users want. We found cases where providers of high-quality answers refused to take the MBs offered by the question asker(s). Their contributions to the site can be seen as earning RP and as a form of altruism. This finding is line with existing literature on the hybrid model of economic and social motivators of online contribution (e.g., Rafaeli and Raban found that non-monetary incentives such as comments accounted for some variance in participation on Google Answers [9]). The very existence of MB affords users to refuse taking MB and in turn shows their altruism and makes the community more healthy.

MITBBS/MB has four important characteristics that collectively make it significantly different from other much more popular virtual currency schemes. First, MITBBS is a free service and MB does not cost you real money (unlike Q Coins or WoW point cards). Second, MB has practical utilities on the site, e.g., like buying a virtual hat for your avatar. This is juxtaposed to the points in Beehive [5], which are solely indicator of user status. Third, MB is not officially for sale. The rare cases of RMT of MB we found were indeed examples of MB's positive role in sustaining the community. Fourth, MB can be freely transferred between users. This is a key difference between MB and other commodity-based incentive mechanisms such as the point system in Yahoo! Answers and the one in mimir [7]. These schemes were built upon the idea of commodity economy which is based on the assumption of rational human behavior and is driven by price. They were so carefully designed and controlled in the system so as to promote participation while prevent from gaming and misusing the system. To varying degrees, they achieved these goals. However, accuracy and efficiency are not the sole determinants of the success of an online community. Prior work (e.g., [7]) has shown that market mechanisms can reduce the sense of community which is another critical aspect of these systems.

In contrast, the MB system in our study can be seen to promote gift economy which is driven by social relations. One prominent example we focus on in this paper is that unlike those traditional incentive designs, the MB system supports direct exchanges of incentives among users. Besides the material needs that arguably have been well supported by traditional incentive designs (e.g., commodity-based market mechanisms), this system-sanctioned direct incentive exchange between users enables and facilitates user practices that create an interesting gift economy and fulfill users' broader online community needs. Our study reveals evidence that these user-generated practices help sustain and enrich the community by supporting emotional needs (i.e., not always rational) and community culture like RP.

There is a tradeoff between the rigor of system control and the room for user empowerment and creativity. As we have seen in the Web 2.0 movement, user-generated content has revolutionized the Internet, creating many engaging and sticky user experience. We would argue that in contexts where the consequences of gaming or cheating the system are not so detrimental (esp. no real money is at stake), opening up the space for the users may be a better strategy.

In conclusion, we urge incentive and community designers to also consider the gift economy dimension of online community, to experiment with the idea of direct user incentive exchange, and to further open up the incentive space to its users.

REFERENCES

1. Alex Pham. Facebook mulls over adding virtual currency as coin of its social realm. *Los Angeles Times Tech Blog*, 2009. http://latimesblogs.latimes.com/technology/2009/03/facebook-gdc-vi.html.
2. Cheng, R. and Vassileva, J. Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. *User Modeling and User-Adapted Interaction 16*, 3-4 (2006), 321-348.
3. China Internet Network Information Center. *2008 Report on Chinese Online Gamers (in Chinese)*. 2008. www.cnnic.cn/uploadfiles/pdf/2009/3/24/142752.pdf
4. Comaroff, J. and Comaroff, J. *Ethnography and the Historical Imagination (Studies in the Ethnographic Imagination)*. {Westview Press}, 1992.
5. Farzan, R., DiMicco, J.M., Millen, D.R., Dugan, C., Geyer, W., and Brownholtz, E.A. Results from

deploying a participation incentive mechanism within the enterprise. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM (2008), 563-572.

6. Harper, F.M., Raban, D., Rafaeli, S., and Konstan, J.A. Predictors of answer quality in online Q&A sites. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM (2008), 865-874.
7. Hsieh, G. and Counts, S. mimir: a market-based real-time question and answer service. *Proceedings of the 27th international conference on Human factors in computing systems*, ACM (2009), 769-778.
8. Hsieh, G., Kraut, R., Hudson, S.E., and Weber, R. Can markets help?: applying market mechanisms to improve synchronous communication. *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, ACM (2008), 535-544.
9. Rafaeli, S., Raban, D.R., and Ravid, G. How social motivation enhances economic activity and incentives in the Google Answers knowledge sharing market. *International Journal of Knowledge and Learning 3*, 1 (2007), 1 - 11.
10. Wenjing, X. Virtual space, real identity: Exploring cultural identity of Chinese Diaspora in virtual community. *Telematics and Informatics 22*, 4 (2005), 395-404.

Leveraging PBL and Game to Redesign an Introductory Computer Applications Course

Scott J. Warren
University of North Texas
Department of LearningTechnologies
3940 N. Elm, Suite G150
940-369-7489
Scott.warren@unt.edu

ABSTRACT

The purpose of this paper is to discuss one instructional design that leverages problem-based learning and game structures as a means of developing innovative higher education courses for students as responsive, lived experiences. This paper reviews a curricular redesign that stemmed from the evaluation of an introductory course in computer applications that had high drop, failure, and withdrawal (D/F/W) rates. Interviews with students and faculty in this course revealed that students were not engaged with, motivated by, or satisfied with the instructional methods, which were often frustrating and difficult to navigate. Using data collected from students and faculty, we describe the full redesign of the course, which included ill-structured problems for students to solve, multiple forms of learning assessment, and a contextual framing stemming from a digital, alternate reality game design. When comparing the new design to the original, the first iteration research indicated decreased failure rates, increased achievement on standardized assessments, and a range of individual student experiences from high praise of the design to some disappointment.

Categories and Subject Descriptors

K.3.2 [**Computer and Information Science Education**]: Literacy, *computer literacy.*

General Terms

Performance, Design, Experimentation, Human Factors, Theory,

Keywords

Blended learning; Course redesign; Game; Communicative Action; Problem-based learning

1. INTRODUCTION

The original computer applications course was designed to teach the basics of computer parts, Internet use, online security, and using Microsoft Office™ implemented an Adobe Flash™-based computer assisted instruction program called SAMS 2003 Computer Literacy™ [5].

An analysis of course materials and notes from interviews revealed several issues that led to dissatisfaction with the existing curriculum and instructional methods used in the course. The design team identified the following problems:

1. *Functions in one program were not linked to others.* Understanding how the programs can be used in a complementary fashion is an important objective that wasn't addressed in the existing curriculum. Students found learning the same skills in four different programs boring and repetitive, since many students were already familiar with these basic actions from previous computer applications courses in high school.

2. *Computer-based assessments and instruction were too rigid.* There are commonly three to five ways to perform an action in a program; however, the program often recognized only one. In some instances, the practice exercise asked a student to complete a task in one manner, while the assessment compelled another.

3. *Applications of knowledge were decontextualized and unrealistic.* Based on students' current life experiences, the applications of learning expected in both the training practice and exams don't fit well with an undergraduate's life experiences. Students saw little relevance between course content and their future work. They also found the constant drill and practice tedious.

4. *Computer-based instruction provided weak feedback.* The system provided weak feedback during both training and often none during exams. The amount of feedback did not increase or decrease dependent on how many correct or incorrect answers students provided for a specific objective.

2. THE REDESIGNED COURSE

Based on this analysis, the design team determined that the following measures be taken to address the underlying problems:

1. *The number of discrete learning objectives should be revised from 750 to 150.* Given the length of the course, the sheer number of objectives was overwhelming to students, and should be collapsed to eliminate redundancies from unit to unit. For example, the objectives "The learner will be able to open an MS Word document" and "The learner will be able to open an Excel spreadsheet" should be consolidated to read, "The learner will be able to open documents within MS *Office™*."

2. *Requirements for the course should be stable across sections and semesters, but revised yearly to incorporate innovations in the field.*
With rapid changes in information technology and differing needs of students, course requirements should be updated regularly. An examination of state technology requirements for K-12 learning should take place yearly to ensure that the course does not simply re-teaching the same concepts students learned in high school.

3. *The course should be centered on larger learning projects and problem solving using the software, not around disembodied learning tasks.*
The nature of the computer applications introduced in the course readily lends them to a project-based or contextual learning approach. To better engage students with these tools in the manner for which they're intended, the learning tasks should leverage them as a means to solve an ill-structured problem, design a project, or effectively communicate ideas to others. Development of appropriate, rubric-based assessments rather than multiple-choice tests is also warranted.

The university's retention goals, the research literature, and analysis of the existing course supported a redesign using problem-based learning methods. Furthermore, the use of story-like scenarios typical of problem-based learning (PBL) [3] is a prominent element in digital games, and media products known to engage players for hours on end.

However, given the challenge and cost of designing an immersive game world, alternative media that leverages that both narrative plot and the requisite learning scaffolds to facilitate learning is necessary. One such alternative is to embed game activities and resources in a variety of media, distributed across the Internet using Alternate Reality Game (ARG) structures [2, 4], rather than a fully integrated, stand-alone product. This approach maximizes resources, such as MySpace, generic web logs, Podcasts, YouTube™, and the three-dimensional digital environment of Linden Labs' *Second Life*™.

Students work in small teams of 3-4 students to solve problems posed by fictional clients, similar to those they may encounter in a video game, that last for two weeks during which time they explore the resources provided to them or uncover more as they play they game. This helps create an open system of resource distribution that authentically mirrors the contexts to which learners will transfer the skills and knowledge once they are done with college and working in the real world. This concept also allows designers to exploit many free online resources while merging them with PBL methods to give learners a situated, coherent narrative to contextualize their learning experience while concurrently providing cognitive scaffolds from which to retrieve knowledge and skills necessary to their future work and learning. In order for students to cognitively transition from an acquisition model to knowledge construction requires curricular and instructional innovation.

2.1 Student Learning Outcomes and Assessment

As noted earlier, student learning outcomes (SLO) were reduced from more than 750 separate outcomes/learning objectives to 150, mainly by eliminating those that were repetitive from a single software program to the next. The objectives that emerged may be found here. These outcomes were only those specifically related to surface-level learning related to either memorization of facts or application of skills based on rote memorization.

2.1.1 Assessment

As with any new instructional methodology, it is as important to discover the specific answers as to why it was more effective than another method as it is to learn that it is effective at modifying behavior, improving achievement, or engaging learners in reflection associated with improvements in critical thinking. This research study employed both qualitative methods to examine student experience as well as a quasi-experimental, posttest comparison design to measure the effect of a digital game-based, problem-based curriculum in a hybrid course on student achievement. The posttest questions used the shared learning objectives for each section of the course.

The participants were quasi-randomly selected as they self-selected to participate in each section with no prior knowledge by the researchers. These participants had no prior knowledge of the instructional style of the course prior to enrolling other than knowing that the comparison section is a hybrid course which meets only partly in a classroom, that the section meets entirely online, or another section meets entirely face-to-face each week.

The research of the quantitative questions differed dependent on the outcomes sought by the questions. In the instance of the achievement questions, students have been randomly assigned to a condition dependent upon the section of the course that they signed up to take with no prior knowledge of the research questions and with no influence by the researchers. Two sections engaged in the existing 1.) online and 2.) face-to-face sections acted as the comparison groups. A third section, developed specifically as a hybrid course with six total face-to-face meetings combined with learning tasks and activities to be completed online using open-source and no-cost Web 2.0 tools. This third section acted as the treatment condition. The following were our main research questions linked to overall goals of the course, including achievement:

> *Question 1.* Can the use of a game-driven, problem-based learning curriculum for post-secondary learners that leverages existing and developed distributed learning resources should improve the achievement of learners at a statistically significant level more than those learners in the existing drill and practice-based course?

In order to address this question, students in the two conditions engaged with either the treatment or comparison curricula over the course of the semester. At two points during the semester (midterm and final), each group completed two exams based on the same learning objectives shared by each curriculum.

> *Question 2.* Can the use of a game-driven, problem-based learning curriculum for post-secondary learners that leverages existing and developed distributed learning resources improve the level of satisfaction expressed by learners more than those learners in the existing drill and practice-based course at a statistically significant level?

This question was addressed by providing students with a survey

of learner satisfaction with the overall course, the means of instruction, means of assessment, and learning activities similar to the means of course quality assessment that will be provided by the College of Education. However, while summative, this survey was more in-depth and asked specific questions about the delivery of instruction by the game and online systems, attitude toward instructor, attitude towards instructional style, self-report of instructional style, attitude towards individual learning components and activities, attitude towards peers, as well as general satisfaction with the course.

2.2 Pedagogical approach

Students in this class met face-to-face four to ten times during the semester, depending on instructor and student preference, to practice the basic skills needed to complete coursework. The remainder of the time, the class met in small groups online in the *Second Life™* digital environment, which is free to download and enter. Visual imagery and audio provide information, tasks to complete, and a larger narrative structure within which students may situate their learning. Rather than listening to lectures and taking large multiple-choice tests, students will hone their technology skills by solving ill-structured problems that they encounter in the 3-D environment. They worked in small groups, using productivity tools to develop products that solve posed problems, and take part in support one another using a complimentary courseware tool called *Moodle.* The goal of this instruction was to provide students with a general set of skills that will allow them to use any word processor, spreadsheet program, or presentation tool and adapt to new versions readily. Problems were contextualized within a larger narrative structure that takes students along through a linear story within which their understanding was be situated. Game structures such as an overarching conflict, objectives to complete, clues to seek out and interpret, and feedback from the 3-D system were expected to increase student engagement with the learning of basic computer skills and knowledge.

Both problem-based learning (PBL) and alternate reality game structures were used to redesign the computer applications course. Rather than listen to lectures, complete practice exercises, and take frequent multiple-choice tests, students hone their technology skills by solving a series of ill-structured problems posed by fictional clients using the very tools they are expected to learn. Students work on each task or problem in small groups of two or three, using a variety of productivity and communication tools. The redesign made use of a hybrid or blended learning format. Face-to-face class time was dedicated to delivering instruction and to facilitate group problem solving. Online resources, support, and collaboration tools were also provided through the free courseware platform, *Moodle.* However, students were encouraged to make use of whatever productivity and communication tools best fit the dynamic of their groups. Emphasis was placed on communicating with peers, in class and online, to develop viable and deliverable solutions, rather than enforcing conformity to a specific version of a designated proprietary software program. The goal of this instruction was to provide students with a general set of skills that would allow them to use any word processor, spreadsheet program, or presentation tool and adapt to new versions readily. It also addressed some of the issues cited previously that frequently accompany online learning and digital collaboration, compelling students to negotiate solutions to issues of

accessibility, software compatibility, and file management in their own teams. For example, if one team member could not afford the latest version of *Office™*, the team might use Sun's *Open Office™* or Google Docs™. These broadened objectives were expected to better prepare students for their future world of work.

2.2.1 Alternate Reality Game course structure

The Door ARG was designed with a two-tiered narrative structure that framed course activities and provided the context for problem solving. The first tier of this narrative engaged students with fictional clients who "hired" student teams to complete authentic tasks — a problem-based narrative approach. The second tier engaged students in game structures that included puzzles, codes, and ciphers that must be solved, retrieved or used correctly in order to gain access to materials, information, and resources that provide additional scaffolding and narrative support to the first tier learning tasks. In essence, each of the clients and characters in the six, PBL scenarios had alternate personas, hidden beneath their client identities, and all of them were embroiled in an underlying conflict with each other as well as the unsuspecting student players. Within the top-level story of *The Door*, students are asked by "clients" to solve complex, ill-structured problems that require them to use all the major components of Microsoft *Office*™. The problems students faced ranged in complexity. In one instance, students had to provide directions to an inept old coach and gym teacher for how to construct a properly functioning grade book spreadsheet in order to allow him to keep his job at a middle school. In another instance, students develop an improved web site for a local nightclub that included appropriate use of basic color theory and space usage.

At the same time, clues appear indicating that a software program called the Autumnal Equinox Firewall has disappeared which may have dire consequences for both the students and the world. Through these clues, the second tier of the story is revealed. The Puppet Master character of the game, Hester, offers students rewards for locating relevant game information. Further, she notes that additional resources to improve their problem solutions will be revealed if they obtain these rewards, such as a video that students can locate in YouTube™ (link here) if they put together a web address correctly. In this way, the Puppet Master, played by the instructor, provides soft scaffolding and additional resources for students who may be struggling either to solve the ill-structured problems or locate game resources. Game characters also act as gatekeepers, judging the quality of student solutions and preventing them from moving to the next problem until the last has been adequately addressed. As students move through the story at both levels, clues and minor puzzles are revealed. If students are successful at piecing this information together, they may discover that the clients are intended to be the ancient Greek gods seeking to reclaim followers and power by harnessing the power of the Internet, a power these students seek to understand.

Figure. *Distributed resources, clues, and communication tools for the ARG.*

Visual imagery and audio were used to provide objective information about the ill-structured learning tasks and to spur communications among students, instructors, and game characters. Further, the two-tiered narrative framing provided the means for students to situate their learning in a more meaningful and engaging context [6]. It also encouraged students to interrogate the inconsistencies between the two plotlines and was leveraged to challenge students to rethink their surface-level understandings of what was presented to them by the game.

3. RESULTS

Given that this course redesign stemmed from a university initiative aimed at improving retention, satisfaction, and academic achievement, research into the redesign employed multiple methods: qualitative and quantitative. The quantitative measures included examination of drop, failure, and withdrawal rates, analysis of posttest achievement scores, and a measure of student satisfaction with the course. Qualitative data from semi-structured interviews with students in the pilot implementation was collected and analyzed using constant-comparative methods. This data was triangulated with data from the aforementioned student blogs which were analyzed through computer mediated discourse analysis [1].

3.1 Quantitative findings

The analysis of scores from first, Spring 2007, implementation of the experimental curriculum had mixed, but promising results on measures of retention, satisfaction, and achievement as shown in Table 1.

Table 1. Quantitative results for student retention, satisfaction, and achievement

	Comparison *n*=57	**Treatment** *n*=32	**Differences**
Retention (% DFW)	21.05%	12.50%	- 8.55%
Drops	1	2	
Failures	2	0	
Withdrawals	9	2	
Satisfaction	3.64	4.2	alpha=.05, z(6)=6.86, p=1.64
Achievement	M=78.83	M=85.96	t=3.90, crit=1.67

The results indicate an 8.55% difference in the percent of students who dropped, failed, or withdrew between the comparison course and the treatment. Moreover, satisfaction with the redesigned course, as gauged by the five item college course

evaluation which is the measure used at the university, was statistically significantly higher than in sections using the existing course design. Finally, student achievement, as measured by posttest in both groups, and compared using a two-sample t-test assuming unequal variances showed greater improvement in the treatment group the comparison group.

3.2 Qualitative findings

Although student satisfaction was higher in the redesigned course, qualitative data collected through interviews with students tells a slightly different story. Interviews indicated that students gained a number of insights and understandings from the experience. These included:

1. An appreciation for how the technology skills gained in the course applied to the world of work and would impact their future.
2. An understanding of the significant role that interpersonal communications play in learning and in career success.
3. A sense of empowerment fostered first by access to resources and later by development of the knowledge and skills to become resourceful
4. An increased willingness to play, explore, and experiment with tools, content, and processes that points to potential lifelong learning [7].

4. DISCUSSION AND CONCLUSION

The course continues to be taught in different forms each semester. These differences stem from data collected each semester from students and faculty teaching the course. Upon reviewing the qualitative data collected over the course of the last three years to contextualize the quantitative findings, communication problems among students in their groups was paramount, indicating that this is a skill entering freshman substantially lack. This also had roots in the design of the course that required high levels of communication using digital tools as a means of completing work and asking questions. Problems of communication were further exacerbated by the size of some of the groups that ranged from three to five, depending on student choice. Finally, a lack of student experience with group communication and problem solving that they brought with them to the course from high school and other undergraduate courses further complicated matters. While the problem-based learning component was challenging for some students, they made clear linkages between problems they were solving in the course and those they would have to solve in the future. However, while many students reported disliking working in groups to complete tasks, they did recognize its necessity in their future careers and that excellent interpersonal communications skills are necessary for their future success. The next iteration of this course will undergo a full redesign in order to specifically target each of these communicative goals and allow for the evaluation of student and instructor success at reaching them. We will also redesign the problem-based aspect of the course and the game aspect so that students must complete the game in order to successfully complete the course, which we hope will result in even higher student satisfaction rates amongst students.

While the instructional methods have yielded some mixed results stemming from the use of problem-based learning and its accompanying reliance on students to self-organize and

solve ill-structured problems, it has provided a wealth of data and results related to improving student experience with experimental and innovative instructional methods that push towards the edges of what our students are capable of as they enter college. Overall, the results of the research related to this course redesign leave the researchers hopeful that the hybrid course and use of an ARG to frame the problem-based learning tasks and group interactions were responsible for the improved test scores. While the statistical findings of this study were mixed in terms of student retention, the increased posttest and ratings scores are promising and will be followed up with additional studies on future iterations of the course.

However, the process was not without difficulty. There have been many challenges to the innovation from student, instructor, and technology arenas, causing this course to go through eight iterations over the course of the past three years in response to student and instructor realities [7]. Taking the innovation from a single section of 32 students to seven sections and more than 200 students while coordinating five or more instructors with wildly different pedagogical and epistemological views has been difficult and extremely time consuming. Developing additional materials on a weekly and monthly basis in response to instructor needs and questions or simply answering questions and assuaging fears sometimes takes as much as 25 hours a week in addition to monthly afternoon meetings. While the adjunct faculty teaching the course have been supportive, they have complained often of the increased workload required by the course in terms of the amount of feedback and grading they are required to provide, the level of communication they must maintain with their 32-64 students to respond to questions by e-mail, phone, or other digital structure, and the general lack of student willingness to read directions and do a reasonable level of work in the course.

Students were often not prepared by high school for the innovative curricula they should expect in college. Therefore, we must work more closely with professionals in K-12 settings and at state agencies to better prepare students and instructors for the critical thinking and creative tasks they should expect in college and beyond. Developing innovative curricula in high school that both targets acquisition of knowledge and skills while challenging students to solve ill-structured problems and be prepared for self-direction, ambiguity, and critical thinking will require that instructional designers and educators on both sides of the secondary/post-secondary line challenge the fundamental objectives of schooling. By doing so, we prepare our students for the future world of work in meaningful ways that allow them to be successful in the 21st century Conceptual Age economy rather than be left behind because they are not prepared to adapt to the rapidly changing world around them.

5. REFERENCES

[1] Herring, S. C. 2004. Computer-mediated discourse analysis: An approach to researching online behavior. In Designing virtual communities in the service of learning. Barab, S. A., Kling, R. & Gray, J. H.. Eds. Cambridge University Press. Cambridge, UK. 338-376.

[2] Martin, A., & Chatfield, T. 2006. Alternate Reality Games White Paper - IGDA ARG SIG. International Game Developers Association. Mt. Royal, New Jersey.

[3] Savery, J. R., & Duffy, T. M. 1995. Problem based learning: An instructional model and its constructivist framework. Educational Technology, 35, 31-38.

[4] Terdiman, D. 2008. 'The Lost Ring' ARG players discover 'lost' Canadian sport. *CNET News.com*.

[5] Thomson Course Technology. 2007. Building the Bridge to Better Microsoft Office Instruction: Report on a Survey. Technical report. Thomson Course Technology/Cengage.

[6] Warren, S., Stein, R., Dondlinger, M., & Barab, S. 2009. A look inside a design process: Blending instructional design and game principles to target writing skills. J. of Educational Computing Research. 40, 3. 295-322.

[7] Warren, S. J., Dondlinger, M. J., & Whitworth, C. 2008. Power, Play and PBL in Postsecondary Learning: Leveraging Design Models, Emerging Technologies, and Game Elements to Transform Large Group Instruction. Paper presented at the American Educational Research Association Annual Meeting. AERA '08 (New York, New York, March 24-28, 2008.) AERA Press.

Name Matters: Taxonomic Name Recognition (TNR) in Biodiversity Heritage Library (BHL)

Qin Wei*
Graduate School of Library and Information Science
University of Illinois
501 E Daniel St
Champaign, IL, USA
qinwei2@illinois.edu

P. Bryan Heidorn
School of Information Resources and Library Science
University of Arizona
1515 East First St
Tucson AZ, USA
heidorn@email.arizona.edu

Chris Freeland
Missouri Botanical Garden
4311 Shaw Blvd
St. Louis, MO, USA
chris.freeland@mobot.org

ABSTRACT

Taxonomic Name Recognition is prerequisite for more advanced processing and mining of full-text taxonomic literatures. This paper investigates three issues of current TNR tools in detail: (1) The difficulties and methods used in TNRs. (2) The performance of Optical Character Recognition (OCR) and TNR tools by samples from Biodiversity Heritage Library (BHL). (3) The methods for potential improvement. We found that the performances of current TNR techniques need to be improved. A detailed error analysis reveals that sublanguage characteristics account for much of the error. A preliminary experiment using NaiveBayes (NB) models shows the potential of using machine learning (ML) in TNR.

Categories and Subject Descriptors

H.3.7 [**Digital Libraries**]: Systems Issues, User Issues; I.2.7 [**Systems Issues, User Issues**]: Natural Language Processing

General Terms

Algorithms, Design, Performance, Experimentation, Languages

Keywords

Taxonomic Name Recognition, TNR, biodiversity informatics, Machine Learning, Digital Libraries, Information Retrieval

1. BACKGROUND

Digitization of library materials has become a global trend especially for biodiversity informatics such as the BHL (`http://www.biodiversitylibrary.org/`) project. BHL has been funded through a sub-award from the Encyclopedia of Life to digitize more than 60 million pages of legacy scientific literature within 5 years. 25,995,854 pages are available to date via the BHL Portal (updated 11/15/2009). An important aspect of BHL is the incorporation of "taxonomic intelligence" provided by the Universal Biological Indexer and Organizer (`uBio:http://www.ubio.org`) to automatically identify taxonomic names. The image files created by high-resolution scanners are processed through ABBY FineReader or PrimeReader (OCR softwares) to create text files. Those text files are then submitted to uBio's TaxonFinder web service to identify the candidate names. All candidate name strings are compared to NameBank, uBio's repository of about 10.7 million scientific names. When a match is made, the verified name string is then made available for search and display in the BHL portal.

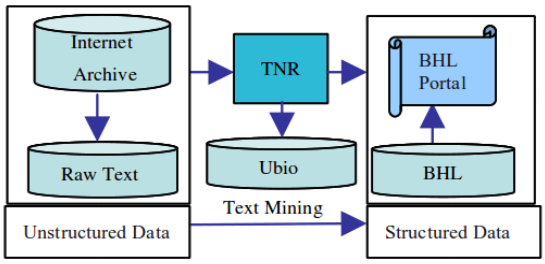


Figure 1: BHL architecture

Figure 1 presents current BHL architecture. The ultimate goal of the BHL is to build an intelligent user-driven digital library that provides the most authoritative information on all species and the means to navigate and analyze the information.

The paper is organized as follows. Section 2, introduces the methods used, and details why TNR is difficult. Section 3 presents the experimental design and characteristics of the BHL collection. Section 4 details the performances of OCR. Section 5, shows the performances of TNRs, an in-depth error analysis and methods for potential improvement. Section 6 presents the discussions and future work.

*Corresponding author.

Table 1: The methods used in TNR

Method	How	Example
Dictionary Lookup	Compare the target string to the strings in a dictionary	FAT [7]
Rule-Based	Using domain knowledge to construct rules	TaxonFinder [3]
Machine-Learning	Using corpus information to make decisions	MARTT [1] and Herbis [2]

2. INTRODUCTION

2.1 TNR Methods

Taxonomic Name Recognition could be regarded as a subtask of Named Entity Recognition (NER). TNR "has been developed to exploit the linguistic and contextual nature of taxonomic names, as dictated by Linnaean rules used for most organism scientific names since 1754" [6]. Many methods used in NER are adopted in TNR. The most common ones are dictionary lookup, grammars and rules matching, and machine learning, as summarized in table 1. Currently most real life applications are some combinations of the three methods while they might take advantage of one method more than the others. The two TNR tools evaluated here are TaxonFinder (`http://www.ubio.org/index.php?pagename=soap_methods/taxonFinder`) and FAT (Find All Taxonomic names. `http://idaho.ipd.uni-karlsruhe.de/GoldenGATE/`), both of which adopted the combination approach. TaxonFinder relies more on rules while FAT focuses more on dictionary lookup. The two are selected because they are the most widely used tools within the biodiversity community.

2.2 Difficulties of TNR

The main challenge of the TNR task and the requirement for an effective algorithm to perform this task depend on the type and degree of name variations in the collection to which the matching algorithm is applied. The name variations can be divided into three types: different kinds of literatures (language, genre, age), taxonomic naming variation and OCR errors. The variations in this research are listed as following:

1. BHL collection is a typical biodiversity collection that contains a huge volume of diverse literatures. The varieties include multi-language, a long time span (from 1500 to present), multi-discipline, multi-genre (journals and books) and so on.

2. Naming variation is ubiquitous in all kinds of taxonomic literatures. The variations increase the difficulties for any kind of automatic text processing. An informal analysis was conducted by the author and several main categories of naming variations were identified: with/without Species Author, genus abbreviation, Species author and genus abbreviation, invalid strings following correct name and the combinations of them. But there are also many exceptions, which are not listed here.

   (a) Variation because of author string: e.g. Cytisus supinus and Cytisus supinus Pimpinella, Cetraria aculeata and Cetraria aculeata (Ehrl.), Smelophyllum capense and Smelophyllum capense Rdlkf;

   (b) Variation because of Genus abbreviation: e.g. Amoora speciosa and A. speciosa;

   (c) Variation because of Genus abbreviation and author: e.g. Baeomyces intermedia and intermedia (Del.), Cladonia fimbriata and Cl. fimbriata Hffm (Note the Genus abbreviation has 2 letters), Cladonia pungens and Cl. pungens (Ach.), Durio carinatus and D. carinatus Mart;

   (d) Variation because of invalid characters following correct name: e.g. Orobus albus and Carduus mollis ("Orobus albus" is right, but "Carduus mollis" is not a valid author or name);

   (e) Variation because of c & d: e. g. Parmelia conspersa and P. conspersa Ach - Usque; "Parmelia conspersa" and "P. conspersa" shoud match. "Parmelia conspersa" and "P. conserpersa Ach." should match. However, "Parmelia conspersa" & "P. conserpersa Ach. - Usque" should not match. "-Usque" is not a valid name.

3. The most important factor is introduced by OCR errors. Since we are automatically transforming the image files into text files, errors are introduced at the same time. Although we are able to identify the most frequent patterns of OCR errors as presented in Section 4, generally speaking, the errors are unpredictable in a sense that the error patterns in real texts are irregular.

However, it is those difficulties in TNR make it different from other NER tasks and interesting to information researchers.

3. EXPERIMENTAL DESIGN

Three related questions are going to be answered by the following analysis: (1) Performances of OCR and TNRs in BHL. (2) Error analysis. (3) Candidate methods for improvements and the expected performances.

3.1 Experiment Procedures

For answering the first two analyses, first we need to construct the ground truth from the sample pages. We sent the pages to 14 volunteer biologists recruited at the beginning of the project along with a excel spreadsheet. The procedure began with their manually identification of all valid names in each page. The spreadsheet includes three columns: pageid (BHL unique identifier for a page), name_as_printed, names_as_OCRed. Name_as_printed and name_as_OCRed record the characters represent the names as printed and in OCRed text. The name_as_printed is served as the ground fact for the following discussions. The OCRed texts are used to evaluate the OCR performance. The names identified by TaxonFinder were retrieved from the BHL portal. The results include pageid and names identified. Software used for testing FAT is called GoldenGate (`http://idaho.ipd.uni-karlsruhe.de/GoldenGATE`). The version is 2008.03.25. 20.30. The results from FAT include the same fields: pageid and names identified.

Table 2: Characteristics of the sample

Number of Pages	392
Average Number of Words per Page	446.8
Average Number of Names per Page	7.7
Total Number of Names	3003

For answering the third question, a NaiveBayse (NB) classifier is implemented since NB is usually used as the baseline classifier in machine learning (ML). The toolkit used in this experiment is called WEKA version 3.4 (`http://www.cs.waikato.ac.nz/ml/weka/`). A 5-fold cross validation method is then used to evaluate the performance of NB.

Evaluation measures used in this study are standard Information Retrieval (IR) evaluation measures: precision, recall and F-score (detailed information about the measures could be found in Salton, 1971 [5]).

3.2 Sample Characteristics

We randomly selected 392 pages from the BHL database that contained 4,843,619 pages at the beginning of our project. Table 2 shows some characteristics of the sample. We denote a word to be any sequence of one or more letters that begin and end with a punctuation or space. We can see that the literatures are rich in names.

Meanwhile, we categorize the pages into three types: index pages, sublanguage pages and regular pages. Index pages are those pages that do not have grammars. Generally speaking, they include a list taxonomic names with/without page number. Sublanguage pages contain the most important taxonomic information. Here is an example: "Plants terrestrial, on rock, or rarely epiphytic. Stems erect or nearly erect, rarely long-creeping, scaly."[1] Sublanguage is different from natural language (or complete language, such as English or Chinese) from the perspective of vocabulary, grammar and more importantly, how it carries knowledge. In sublanguage, not only words but also grammars carry meanings. Regular pages are the pages include complete sentences that could be processed by regular NLP techniques.

These three different page types contain very different information. For index pages, taxonomic names appear intensively. Taxonomic descriptions are usually in Sublanguage pages while they contain fewer names than index pages but more than Regular pages. And those descriptions contain morphological information of species that are of importance to biologist. Regular pages may contain any kinds of information but fewer names.

Within our sample of 392 pages, 25 are index pages, 110 are sublanguage pages and 257 are regular pages.

4. OCR

4.1 OCR Performances

OCR, transforming the images to text files, is very important since the results of OCR are where the TNRs are going to be applied. Since the two TNR tools use morphological features to identify the name, we consider the OCR a failure if one or more letters of the generated word are wrong including those in wrong case. For example, one of the rules for matching names could be: Genus name is capitalized and is probably followed by lowercase species and subspecies names. Thus, if a capitalized word were not correctly recognized, the matching process would fail.

[1]Flora of North America online: `http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=10072`

Table 3: Overall OCR performance

Total	Wrong OCR	Error Rate
3003	1056	35.16%

Table 4: Frequent OCR error patterns in BHL

1	Insert Space	8	n->v
2	Omit Space	9	l->i
3	e->c	10	r->i
4	u->i	1	u->ii
5	u->n	12	h->l
6	i->l	13	h->ii
7	c->e	14	e->o

The table shows among the OCRed text of the 3003 valid names, 1056 of them contains at least one wrong character. It's worth mentioning that the performances of the OCR might be very different comparing to other types of text. Our evaluation collection is multi-language and older compared to other collections used in similar studies (e.g. [4] Rice, Kanai, and Narker, 1993). And the target for evaluation is limited to the name string that also makes a difference. We are also able to identify the top OCR error patterns in our sample as listed in table 4.

4.2 Performances On Different Languages

Our sample includes 242 English pages and 150 non-English pages. The precision for English and non-English pages are 64.78% and 64.04% respectively. A student t-test gets the p-value of 0.3333 which is not significant which means there is no significant different between the OCR performance on different languages. The result is reasonable since our target is only limited to taxonomic name strings. And name strings in taxonomic literatures tend to be Latinized in most circumstances where the language of the rest text in the page might be German, Italian or even Chinese.

4.3 Performances On Different Page Types

The OCR performances over different page types are 62.41%, 62.77% and 68.29% respectively for Index, Sublanguage, Regular pages. Several t-tests at 5% level show that the OCR performance on Regular pages is significantly better than the other two types of pages. But the performance difference is not significant between Index pages and Sublanguage pages.

5. TNR

5.1 TNR Performences

In digitization projects such as BHL the algorithms must also be able to find names even if they have OCR errors. So our evaluation included name strings that were identifiable

Table 5: Performances of TaxonFinder and FAT

With_OCR_Error	TaxonFinder	FAT
No. of Names (identified by biologist)	1696	1937
No. of Names Found by algorithms	1540	1603 %
Correct	621 %	452 %
Precision	40.32%	28.20%
Recall	36.62%	23.34%
F-score	38.47%	25.77%
With_OCR_Error	TaxonFinder	FAT
No. of Names (identified by biologist)	2610	3003
No. of Names Found by algorithms	1540	1603
Correct	674	517
Precision	43.77%	32.25%
Recall	25.82%	17.21%
F-score	34.80%	24.73%

Table 6: TNR performances on different page types

TaxonFinder	Precision	Recall	F-score
Index	32.82 %	17.89 %	25.36 %
Sublanguage	39.71 %	24.86 %	32.28 %
Regular	60.67 %	36.11 %	48.39 %
FAT	Precision	Recall	F-score
Index	58.01 %	20.94 %	39.47 %
Sublanguage	17.42 %	12.72 %	15.07 %
Regular	35.35 %	18.24 %	26.80 %

by humans even when they had OCR errors. Both TaxonFinder and FAT employ some forms of fuzzy matching that tried to addresses this problem. For example, Cardnus mollis is a valid name that could be found in page (`http://www.biodiversitylibrary.org/page/22001`). The OCR output Carduus mollis where 'n' was changed to 'u'. But TaxonFinder is able to correctly find Carduus mollis while the string is not confirmed by NameBank. Therefore, it is not shown in the portal. Here we present the performances of both algorithms under the situations with or without OCR errors.

Three t-tests at 5% level show that TaxonFinder is significantly better than FAT in precision, recall and F-score in both scenarios. But even TaxonFinder only achieved an F-score of 34.90%, which is still relatively low for an efficient retrieval. It means among the all names in the literatures, the TNR is only able to identify a quarter of them while leaving out the majority of the names. And among those found names, at least half of them are invalid names.

5.2 Performances on Different Page Types

Table 6 presents the performances of TaxonFinder and FAT in different page types. Both algorithms have significantly different performances over different page types while performing better in different types. TaxonFinder has its best performance in Regular pages and worst performance in Index pages. However, FAT has its best performance in Index pages and worst performance in Sublanguage pages.

Table 7: Page type classification confusion matrix

Confusion Matrix	Index	Sublanguage	Regular
Index	21	3	1
Sublanguage	1	81	28
Regular	1	22	234

5.3 Machine Learning (ML) Approach?

How to improve OCR softwares performance is not the focus of this research. Improving TNR algorithms effectiveness is the main focus. Compared to parsing by dictionary lookup and rules, ML has its own advantages that makes it much more suitable for TNR in OCRed text parsing. Here, we hypothesize that there are two approaches that might lead to the improved performance of TNR.

(1) We can see from table 6, a possibly better tool could combine the result of TaxonFinder and FAT by different page types (use TaxonFinder for Sublanguage and Regular pages, and FAT for Index pages) if page types could be efficiently and effectively identified.

(2) Also, we can see that sublanguage pages have the worst performance for FAT and modest performance for TaxonFinder. We propose that using machine learning to parse sublanguage pages would improve the performance.

5.3.1 Page Type Classification

By using the same sample data, a small-scale page type classification experiment was conducted in order to show the feasibility of combining the results from different page types. NaiveBayes (NB) is selected for this experiment since it is commonly used as the baseline model in various machine-learning tasks. The procedures of selecting features for NB model are explained as follows. Since we are dealing with multi-language pages, the features from a single language might not work well. Instead, we look into the generic features (e.g. morphological features) that exist in a broader range of languages. The features include 1-gram character, 2-gram characters, 3-gram characters, number of words per sentence, and number of sentences per page. We used a NaiveBayes classifier and evaluated it with 5-fold cross validation. The precision is 85.71% and the confusion matrix is shown in table 7.

The performance level we achieved is not trivial. We can see that automatic classification of page types is feasible and could achieve a high performance by combining the results. Despite the performance of the classifier, there are some important aspects are worth mentioning. First, as we can see from the confusion matrix, the main errors are coming from the confusion between sublanguage and regular pages. Part of reason is that some of the pages include both languages, e.g. (`http://biodiversitylibrary.org/page/2496490`) the last paragraph is sublanguage while the other paragraphs are regular language and similar situation in page (`http://biodiversitylibrary.org/page/3050492`). Second, the performance we gained here is the baseline performance that means the performance gained here is the lower bound of the performance we could get by using machine-learning methods. A better classifier could be gained from more carefully selected features and better

Table 8: NaiveBayes performance on text classification

Class	Precision	Recall	F-Score
Name-String	62.60%	20.60%	41.60%

classification models (e.g. Support Vector Machines) that will be our future work.

5.3.2 Sublanguage Pages

Improvement could also be achieved by improving the TNR performances on sublanguage pages. We could see from Table 6, the performances on sublanguage pages is significantly worse than other type of pages for both TaxonFinder and FAT. Supervised learning has been proposed to gain better performance on information extraction from sublanguage text [1]. A small-scale experiment conducted on name string classification using NaiveBayes also showed the potential of using machine learning in this task. The features used are similar with the features used in the page type classification, that include 1-gram character, 2-gram characters, 3-gram characters, Capitalized word or not. We also used a NaiveBayes classifier and evaluated the performance with 5-fold cross validation. The result we achieved is presented in table 8.

We can see the performance we get from an simple implementation of NaiveBayes is an F-score of 41.60% which is encouraging. Despite the performance of the classifier, it is also worth mentioning the following points. The training size on Name strings and non-names in this experiment is very skewed. The size of non-name strings is 20 times larger than name strings size. Skewed training data would lead to a lower performance of machine learning. One possible improvement would be boosting the training size by using Latin taxonomic name dictionaries or the names from NameBank which again will be our future work. Adapting a better classification model such as SVM and more carefully selected features have a great chance of improved performances.

6. DISCUSSION

The performances of OCR and TNRs are presented in section 4 and 5 and the error analysis leads to two proposed methods. We found the large gap between actual and potential performance of taxonomic recognition suggests a possibly fruitful avenue for the improvement of the taxonomic recognition quality. Given the availability of some start-of-the-art named entity recognition methods and the researches done on noisy information retrieval, it is possible to upgrade exiting methods which would substantially narrowing the gap. The characteristics of sublanguage pages and OCR errors made machine-learning methods very attractive. Two potential improvement methods are presented and evaluated in 5. More advanced techniques will be our future work.

7. REFERENCES

[1] H. Cui and P. B. Heidorn. The reusability of induced knowledge for the automatic semantic markup of taxonomic descriptions. *Journal of the American Society for Information Science and Technology*, 58(1):133–149, 2007.

[2] P. B. Heidorn and Q. Wei. Automatic metadata extraction from museum specimen labels. *Dublin Core and Metadata Applications*, 12(2):291–301, September 2008.

[3] D. Koning, N. Sarkar, and T. Moritz. Taxongrab: extracting taxonomic names from text. *Biodiversity Informatics*, (2):79–82, 2005.

[4] S. Rice, J. Kanai, and T. Nartker. An evaluation of ocr accuracy. Technical Report ISRI TR-93-01, University of Nevada, Las Vegas, April 1993.

[5] G. Salton. *The smart retrieval system: experiments in automatic document processing*. Prentice-Hall, 1971.

[6] N. Sarkar. Biodiversity informatics: organizing and linking information across the spectrum of life. *Briefings in Bioinformatics*, 8(5):347–357, 2007.

[7] G. Sautter, K. Böhm, and D. Agosti. A combining approach to find all taxon names (fat) in legacy biosisystematics literature. *Biodiversity informatics*, (3):41–53, 2006.

Innovative Technology in the Classroom: A Live, Real-Time Case Study of Technology Disruptions of the Publishing Industry

Mitchell Weisberg
Sawyer Business School, Suffolk University
Managing Director, Lumen, Inc.
16 Arrowhead Road
Weston, MA
Office: +781-894-9202; cell: +1 781-249-3750
miw3@cornell.edu

ABSTRACT
We are now in the second wave of information technology disruption in the media industry, or more specifically, of digital technology fundamentally transforming our information industries: communications, music, media, publishing and the industries that have grown up around those media. In this paper we will be examining one aspect of the nascent electronic readers (eReaders) disruption of the publishing industry. We will focus on the university textbook (eTextbook) segment. We will examine the impact of eReaders on the university classroom and how they create the opportunities for innovative techniques in the classroom which enrich the students learning experience and provide students with a broader experiential learning environment than was previously possible.

In the Sawyer Business School, Suffolk University, Department of Strategy and International Business, Boston, MA, a section of the strategic management class students are exploring the technology disruption and industry response in the book publishing market. They are focusing particularly on the textbook segment of the market which is being significantly disrupted by the advent and influx of electronic readers and digital textbooks. The class, Management Strategy 429 is the capstone course at the Sawyer Business School. Mitchell Weisberg teaches a section of the course with a focus on business and industry responses to disruptive technologies. Weisberg is bringing innovation to the classroom through the use of digital readers and eTextbooks to create a "Live Case Study" in information technology industry disruption. These innovations enhance the teaching to make this an experiential learning environment and practical, hands-on experience for the management students in the class.

General Terms
Management, Measurement, Documentation, Performance, Design, Economics, Experimentation, Human Factors, Standardization, Theory

Keywords
Innovation, classroom, Kindle, pedagogy, Sony, disruptive technology, information supply chain, eReader, digital book, digital textbook, Amazon

==

1. INTRODUCTION

Over the past centuries and even more so in recent decades we have seen technologies transform industries with profound impacts on society and culture. Examples range from the steam engine transforming an agrarian society to an industrial society and the transistor transforming industrial society into information society. Over the past decade we have observed the transformational impacts of information technologies changing industries: consider the music industry and peer-to-peer technology, or how the communications industry is being impacted by text at the low end, and video at the high end. Each of these transformations has significant social, business and economic impact. The cycle of these technology driven disruptions are becoming more frequent. They are almost part of the daily life in business. These disruptions challenge existing strategies and business models of incumbents and create the opportunities for new strategies for emerging businesses. We are now in the second wave of information technology disruption in the media industry, or more specifically, of digital technology fundamentally transforming our information industries: communications, music, media, publishing and the industries that have grown up around those media. In this paper we will be examining one aspect of the nascent electronic readers (eReaders) disruption of the publishing industry. We will focus on the university textbook (eTextbook) segment. We are examining the impact of eReaders on the university classroom and how they open the opportunities for innovative techniques in the classroom can both enrich the students learning experience and can provide students with a broader experiential learning environment.

2. INTEGRATING CONTENT AND CONTEXT IN THE LEARNING ENVIRONMENT

In the Sawyer Business School, Suffolk University, Department of Strategy and International Business, Boston, MA, one section of the strategic management class students are exploring the technology disruption and industry response in the book publishing market. They are focusing particularly on the textbook segment of the market. This book market is being significantly disrupted by the advent and influx of electronic readers and digital textbooks. Management Strategy 429 is the capstone course at the Sawyer Business School. Mitchell Weisberg teaches one section of the course with a focus on business and industry responses to disruptive technologies. This fall, 2009, the class is examining the impact of digital technology on the textbook publishing industry. Weisberg is bringing innovation to the classroom through the use of digital readers and eTextbooks to create "Live Case Study" in information technology industry disruption. The use of these devices in studying this market has integrated the content of the course with the context. These innovations enhance the teaching to make this an experiential learning environment and practical, hands-on experience for the management students in the class. This is an opportunity that we are providing to senior business students that would elsewhere most likely be available only to MBA students. This model is also applicable to MBA and other graduate business students.

At many business schools, students study business cases of companies from the past that have already resolved a situation; they learn from past successes and failures. At Suffolk University we have jumped the time barrier on studying industry cases. We are diving into the industry disruption and are studying the business impacts and decisions in Real Time. ("Real Time Case Study") There are no answers for the strategy questions we are studying - the companies and the industry (textbook publishing) we are studying are wrestling with them just as we are. And we are sharing our information with them.

The Management Strategy students are studying and experiencing first hand the challenges of developing strategy in an industry that is in the throes of technology disruption and transformation. Senior business students in this capstone class are bringing their experience and collective education to develop competitive strategies for the multiple stakeholders in the textbook publishing industry as class team projects. What is unique about the situation is that when it comes to e-textbooks to students are both the market and the business strategists for that market. As users of textbooks they have the perspective of both the market and the industry. The students have the broader opportunity, much like the anthropologist in the field, to study "themselves". This is a real-life case study; the outcome is yet to be determined. In addition, the value created by this educational experience will be delivered to publishers and electronic reader device makers to possibly influence their strategies and future.

3. INCORPORATING INNOVATION IN THE CLASSROOM

An innovative component to enrich the students' experience is the use of digital readers. To make the situation more of an immersion, these college student teams are using the technology and "being the market" which they are developing strategies to support. One team each has Sony Readers, Amazon Kindles, CourseSmart online textbooks, or paper textbooks. There is an additional team filling the role of "wild card" or entrepreneurs who have the freedom to enter wherever they see the greatest business opportunity in the textbook publishing market. Details of the teams are provided below.

Table 1. Live Case Study Teams and Strategic Positions

Information Dissemination Case Team	Primary Focus and Strategic Issues
Team 1: Publisher	The publisher is facing a lot of uncertainty in this new environment. They must determine what their role will be and how they will adapt their business model and strategy to succeed in this rapidly changing market.
Team 2: Sony – eReader Touch	As a consumer product manufacturer, Sony must wrestle with their vision for the interactive book and textbook market. They must determine how to leverage their technology with partnerships and other types of relationships with providers of information.
Team 3: Amazon – Kindle DX	Amazon is perceived as first to market and has established a leading position with their Kindle reader. Will they try to duplicate Apple's iPod position in this market?
Team 4: CourseSmart – Online Textbook	CourseSmart is a cooperative joint venture between several publishers, all of whom could be perceived as competitors. Will CourseSmart continue as a joint venture or take on a new strategy? What is the future role for CourseSmart in this market?
Team 5: Wild-card or Entrepreneurs	There are many opportunities in this tumultuous market for another player, either entrepreneur or other existing companies (e.g. Apple, Google, Barnes and Noble, etc.) to either carve out a niche or to make a play at becoming a significant participant. What will the entrepreneurs do to capitalize on this disrupted market?

Using the digital devices provides students with the opportunity to engage in understanding the tactical and practical aspects of implementing a strategy that involves technology. They have the opportunity, and many of them fall for the trap, to become enamored with the technology of the devices themselves. At some point during the semester the students temporarily lose sight of the fact that they are trying to create devices (or better still, provide information in a form) that meets or feels and emerging or perceived need in the market population. In resolving this issue, they learn one of the most important lessons for entrepreneurs – it's not just about the technology, it's about how you meet the customer and market needs.

Students are engaged in studying the supply chain of the publishing industry – moving information from the content creation by authors through consolidation, refinement, production, distribution and ultimately to its dissemination in the classroom. They are challenged to wrestle with the information management and information distribution supply chain. Decomposition of the information supply chain is an additional aspect of the course, bringing further benefits. It helps students understand the various ways in which information is transformed and how the value of information is created or changes with each of those transformations. Translating this into a business model by understanding how that value can be captured and distributed engages students in developing an understanding the business and financial aspects of information management. Their analysis provides an opportunity to extend learnings from earlier marketing, finance and entrepreneurship classes by structuring formal research and hard data to support and expand the financial aspects of this industry. The classroom setting provides an additional source of data gathering and analysis. Each team of students is harnessing the "wisdom of the crowd" in class discussions to harvest insights, perspectives and analysis that has been shown to be of similar quality as more expansive and expensive market research. This experiential learning provides an additional tool and capability for information management that students will take into their professional careers.

Teams in the class are competing against each other and against the market on a field where all the rules are changing - profit pools, distribution channels, retail bookstores - there is no "level playing field". What is challenging for the students is that there is no one "winning strategy". Depending on their business goal, there can be more than one "winner"; this could be an eReader/device manufacturer, a publisher, a content provider or a newly defined industry role.

An additional innovative component is the use of Wikis and other technology to remove the "time and space" bounds from the students' learning environment. The students are writing and contributing material on a class Wiki to further the use of interactive digital technologies, extending their ability to collaborate in teams beyond their meeting times. In addition, this Wiki is also open to the publishers, device makers and other industry participants, further extending the student experience into the business environment in real time.

At the end of the semester the students present their strategies along with how their strategy will drive success in their selected segment of the market. The multiple perspectives allow each strategy to be potentially successful with different market segments or approaches within those segments. The publishers and device readers are interested in the results the student are developing. And the students are gaining practical experience working with a real business issue, enhancing their value when they take on these responsibilities after leaving school.

4. ENRICHING THE STUDENT LEARNING ENVIRONMENT

By creating simulated companies, for which the students must create successful strategies, we turned the traditional educational model inside out. We have created an "inverse internship" where the industry takes place in the classroom rather than the students leaving the classroom to go out into companies within the industry. Instead of going out to study these businesses and the market, we have brought the market into the classroom by explicitly taking on the characteristics of the target market for e-textbooks. With only public access to publishers and device makers and the devices themselves, students are challenged to engage in "reverse engineering the existing strategy" of these companies. Formal analysis of the market and the forces at play provides a rich context for understanding the current and emerging business environment which these companies are seeking to address. This formal analysis is tempered by the informal analysis and experience of immersion in that market -- the University classroom.

5. RESEARCH ON PEDAGOGY IN THE CLASS

In addition to a class for the students, this class is also the subject of research on the impacts and value of technology in the classroom. The research was proposed and approved by Suffolk University Institutional Research Board (IRB). Students were randomly assigned to the 5 teams and technologies in this class. Impact on differential learning by the individuals in each team is being assessed by systematic testing on a weekly basis. Weekly quizzes using standard questions of factual material contained in the reading will provide an indication of the reading retention by the students on each team. Quiz questions on analysis of the material provides insight into the students' absorption and ability to use the textbook material to draw conclusions; this data is less likely to be differentiated between groups, but is being tested. A standardized examination given by the school to all students provides a baseline for comparison and/or normalizing the differential quiz scores across the different teams. The data will be evaluated at the end of the semester in order to avoid any bias in teaching or grading. Anecdotal data from classroom discussions will augment the hard data from weekly assessment. Content of team final papers will provide additional insight into the differential learning between teams.

Following the completion of the course, the data will be analyzed comparing learning (i.e. assessment results) between groups. These results will be used to test against a hypothesis of no difference in learning between groups. Additional research questions include:

- Do student learn better with one technology or another
- Do students show greater likelihood to use the devices than to read a textbook
- Which technology do students like better
- What are the benefits shortcomings of the eReader technologies

In addition to the quantitative and qualitative research on pedagogy of informatics, Case Study research is being conducted to produce a case study of the impact of technology disruption on one publisher and their response. This research on the impact and response of industry to technology disruption is not within the scope of this paper. This Case Study will be written up and used in future classes.

6. IMPACT OF E-READERS ON THE UNIVERSITY CLASSROOM

The digital readers have had significant impact on the students and on the classroom experience. For the faculty, the devices create the advantages of ability to make the experience tangible. There is a buzz associated with the use of the devices, creating an energy in the classroom. However, there is also a cost. Formatting between devices, and even within devices is not standard. Graphics and page numbers are different or displaced in the devices, requiring greater navigation descriptions in giving assignments or in referring to material in the texts. There are the additional logistics considerations of managing the devices to ensure their procurement, distribution, recovery. This is complicated by the requirement that the devices must also be available for all students to experience, in addition to those on the designated user teams. For the students, there is a learning curve on the use of the new devices. Many students initially tried to print out the pages rather than use the devices. Over time the devices have gained acceptance and have generated competition within the class.

The course is being repeated in oncoming semesters. Many of these challenges will be resolved from the lessons learned in this past course and the acquisition of a larger inventory of devices. In addition, enhancements may create further enrichment in the eTextbooks. For example, future capabilities may enable dynamic links from the eTextbooks to student and faculty created material in real time, to create a single integrated learning environment. However, the dynamic nature of the technology evolution and its disruptive effects on the publishing industry make it unlikely that the course structure will stabilize into a routine that can be easily replicated and repeated.

Weisberg received funding and support from the publisher of a textbook company to make these devices available to the students. He also has engaged other publishers with interest in supporting this effort who were not providing the text. In addition, he approached the manufacturers of various text readers (e.g. Amazon Kindle, Sony Reader) for additional participation in the classroom innovation and research. Sony offered additional support. Student teams are using the digital readers during the course. The readers will be shared in the last weeks of the class so all students will have an opportunity to try them. Feedback on market behavior, strategic insights and eTextbook evaluations will be provided to the textbook publishing companies and digital reader manufacturers.

7. BENEFITS TO THE PEDAGOGY AND KNOWLEDGE OF INFORMATICS

There are significant benefits for the many stakeholders from this classroom innovation. The benefits to the students include exposure to the leading technologies in a rapidly emerging market. Students also have increased engagement in the industry from first hand interaction with the breadth of participants across the information supply chain. This improves their marketability and increases their preparation and likelihood for near-term employment. The faculty role is enriched through the engagement in a "live case study" in which the boundaries of the classroom are extended into the outside business world. New case study information is revealed in the market through advertising, news reports and other company data. This constantly changing provide new opportunities to give examples and put the meat on the skeleton of theory stays relatively constant but provides new examples ensuring that the course material is constantly refreshed and current. Students are encouraged and rewarded for following the class content in outside media. The university benefits through industry and media exposure regarding its deployment of innovative technology and teaching methodologies, potentially leading to attraction of higher caliber students. Additional benefits to students, faculty and university include:

- Practical, real-world company strategy experience
- Opportunity to consolidate academic experience with real business challenge
- Possibility to impact emerging transitional market and industry
- Exposure to company management of new, industry transforming technologies
- Potentially enhanced learning environment through the technology.

There are benefits for the industry participants of the "Live Case Study" as well. Potential benefits for these industry participants (e.g. publisher, device manufacturer, etc.) may include the identification and clarification of strategic and tactical issues in the university market for electronic textbooks, accompanied by real-time feedback on product and services from actual users in their target market. Their participation in the course also offers an opportunity for them to strengthen their relationship with the university, faculty and students.

Understanding of technology impacts on business and, more broadly, on industry will be a significant benefit of the research and student analysis evolving from the foundation built by this course. The analysis of the publishing industry, or dissemination of intellectual content, is building on the knowledge gained from the music industry which experienced a similar disruption and which is moving into stages of resolution. The supply chain redesign, redistribution of profits, and overall transformation of the industry is rich in furthering models for information management models. This research will be further developed in the ongoing Case Study development and analysis.

8. CONCLUSION

Bringing new information technology and devices such as eReaders into the classroom offers significant benefits to enhancing the teaching of informatics and to extending the knowledge base of informatics. The pedagogical benefits include enhancing the learning experience, enriching the classroom environment and engaging students "real time" in addressing business and industry challenges of addressing disruptive technology. The content and contextual benefits of furthering the knowledge of informatics includes a better understanding and modeling of the business and industry impacts and responses to digital technology disruption. These benefits are ongoing, and are derived from the practice and the research described in this paper. There are physical and pedagogical challenges posed by this classroom innovation, but these will challenges decrease with time and stabilization of the course structure.

Intellectual Diversity in iSchools: Past, Present and Future

Andrea Wiggins
Syracuse University
School of Information Studies
337 Hinds Hall, Syracuse, NY 13244
+1-315-443-2911
awiggins@syr.edu

Steve Sawyer
Syracuse University
School of Information Studies
344 Hinds Hall, Syracuse, NY 13244
+1-315-443-6147
ssawyer@syr.edu

ABSTRACT
We provide evidence and discuss early findings of faculty hiring trends among those involved in the iSchool community. To better understand the intellectual heritage and major influences shaping the development of the individual and collective identities in iSchools, we develop a classification of the intellectual domains of iSchool faculty, and present a brief descriptive analysis of the community's intellectual composition. This analysis builds on work from 2007 and additional data collected in 2009. The discussion focuses on sources of, and trends in, interdisciplinary diversity in the iSchools. We conclude with a short discussion the potential implications of these trends relative to the future development of the iSchool community.

Categories and Subject Descriptors

General Terms

Keywords
iSchool, discipline, faculty hiring, interdisciplinary, computing

1. INTRODUCTION
The iSchools represent an ongoing form of innovation regarding the interdisciplinary pursuit of teaching and research in the converging areas of information, computing and the rules of theses in human and social experience. As seen from their web presence, www.ischools.org, the iSchools present themselves as the paragon of a thriving, heterogeneous, interdisciplinary research community. Commentary and empirical work suggest that iSchools demonstrate a different form of academic focus from near neighbors in the academy such as computer science, information systems, science and technology studies, and communication (to name a few of these nearby intellectual spaces), as demonstrated by the interesting variations on school compositions by academic background [9, 27].

The goal of this paper is to share and discuss a descriptive overview of the intellectual underpinnings and institutional characteristics of the members of the iSchool Caucus. The paper continues with a discussion of the motivation for the research and prior empirical studies of the iSchools. We then present the early results of our current ongoing work. In doing this we provide a classification of iCaucus faculty members' disciplinary heritage, and discuss the evident causes for the community-level composition based on the sources of representation for specific areas of study. Finally, we discuss the implications for the community and directions for future research.

2. MOTIVATION
The emergence of new academic entities is a perennial topic of interest in sociology of science (e.g. [24]). The nature and implications of interdisciplinary research, at the levels of projects and communities, are topics in a variety of scholarly communities (e.g. [16, 17]). Our interest, as members of an iSchool and, thus, participants in the iSchool movement is also pragmatic: what trends can we detect and report regarding hiring and disciplinary structures of the faculties that make up the various iSchools?

2.1 What are iSchools?
Collectively, iSchools engage in a broad range of interdisciplinary research pursuits and offer a variety of courses that integrate studies from applied computer science, design, and library science, among other disciplines. Thematically, the iSchools typically focus on some combination of people, information and technology, across a wide variety of organizational and social contexts. As a result, course offerings at iSchools vary widely in accordance with the variety of degree program offerings.

The initial seeds of the emergence of iSchools appear to be an indirect result of a sea change in LIS programs in the 1980's, when several long-standing American Library Association (ALA) programs closed or ceased to maintain their accreditation. Hildreth and Koenig documented the prevalent survival strategies for LIS schools: merger with a larger partner or expansion into IT-related fields [14]. It comes as little surprise to members of the community that over half of the iSchools are represented as mergers or realignments in this analysis. Two iSchools have been successful mergers; Rutgers incorporated LIS with communications and journalism, and UCLA's information studies program partnered with education. Further, a number of hale LIS programs have been organizationally realigned and aggressively expanded their studies related to information technology; these include Syracuse, Pittsburgh, Drexel, Florida State, Michigan, Washington, Illinois and Indiana. Other iSchools were created new, such as at Penn State and Indiana Informatics, to bring together scholars and expand the host university's presence. Still others, like UC Irvine, Georgia Tech, and Carnegie Mellon, reflect the expanding role of an existing program. More generally, it seems clear that the intellectual background of an academic unit is influenced by the structures and interests of the local university where the iSchool exists, more than a shared common identity with others in the iSchool Caucus.

2.2 Interdisciplinarity requires disciplines
Many have noted that interdisciplinary research is both challenging and increasingly imperative to addressing many

intellectual, social and practical problems [e.g., 17]. Developing a better understanding of the factors that allow interdisciplinary academic endeavors to survive and thrive is in the interest of both the iSchools and to the broader scientific community as a means of insight into cultivating interdisciplinary research.

One attribute of interdisciplinary research is bringing together scholars from different intellectual traditions, with the degree earned by that scholar used as a proxy measure of difference. Faculties with a range of degrees among the members are typically seen as being more interdisciplinary, or at least more multidisciplinary, which allows for the possibility of doing interdisciplinary work. Simply: interdisciplinary scholarship demands having disciplinary variation [e.g., 24].

3. PRIOR WORK

The intellectual composition of an academic unit has traditionally been studied through examination of academic hiring patterns, a recurring topic for research in the sociology of science. These studies are typically focused on prestige out of concern for the potentially detrimental effects of particularistic, rather than universalistic, hiring in the academy (e.g. [2, 3, 4, 6, 15, 20, 21, 22]. Collectively, these studies have shown that in longstanding academic disciplines, changes to the social structure are slow and eventually lead to prestige stratification of the fields. While these studies lay the groundwork for the current work, they represent the concerns of mature academic disciplines in which change is slow to permeate the institutional structure. By contrast, the iSchools form an emergent, loosely coupled academic community. While this academic association has been building for some time – perhaps since the 1970s – the iSchools Caucus was chartered in 2005.

As yet, there is little scholarly research on the iSchool community. The annual iConference, currently in its fifth year, serves in part as a venue for reflection by the members on the efforts of the whole. This venue for community development among members of the iSchools Caucus has generated a few self-reflective studies from the community, but most of these are either largely conceptual or anecdotal, although some represent histories in the making [1, 5, 10, 12, 13, 18, 19, 23, 25]. Little of the discourse focused on the iSchools as a phenomenon is based on empirical data. Recent work demonstrates how errors in sampling for such a small academic community can lead to misrepresentations of the member institutions, particularly for uncertified data [7].

A 2007 study of hiring patterns in the iSchools sought to address some of these issues through empirical research on the relationship of prestige to hiring and identity in this emergent academic community [26]. This research compared the structural characteristics of faculty hiring in iSchools and Computer Science departments. A central finding from that study is repeated here: the disciplinary diversity of the iSchool community was evidenced by 674 faculty PhD degrees in 172 areas of study. While a majority of the faculty received degrees in the categories of computer and information sciences or library science, nearly half of the faculty members completed their doctoral study in other disciplines. This finding regarding the diversity of the iSchool community motivates the current research into the interdisciplinary characteristics of the iSchool faculty.

4. CURRENT WORK

The current study expands on the iSchools hiring research from 2007 and subsequent work [26, 28], but focuses on understanding variation in disciplinary training within community. Our study uses similar data that reflect on faculty composition, employing the notion from [27] that the intellectual identity of the current institution will be based in some measure upon the intellectual heritage of its faculty. While this analysis builds on the 2007 data with additional data from 2009, it is not a longitudinal analysis because faculty change over an elapsed period of two years is not sufficient to make meaningful observations.

4.1 Methods

The population for this study is the faculty of the 21 members of the iSchools Caucus as of January 29, 2009. Analyzing any community necessarily requires purposeful sampling in order to represent the phenomenon of interest. Thus, this population selection excludes those schools which may be self-identified as information schools in name or mission, but which have not joined the iSchools Caucus.

4.1.1 Data collection

The sampling frame was drawn from faculty listings on the web sites of the 21 iSchools. Data collection was done in January of 2009, giving each iSchool time to update their sites for faculty changes for that academic year. Still, some schools had less up-to-date listings of faculty than did others at the time of data collection; while some schools were potentially slightly misrepresented, all such data are subject to this bias due to the inevitable delay between hires and web page updates. These considerations aside, the quality of the sampling frame is still improved over previously available methods.

Faculty roles are variously defined among different schools, and roles such as "lecturer" or "associate in information studies" are not necessarily representative of the long-term intellectual investment in academic expertise that our analysis targets. In addition, professor emeritae are more representative of the prior identity states of a school than its current state. For these reasons, only current full-time professorial faculty were included in the sample; these were identified by their standard academic titles of professor, associate professor, assistant professor, associate dean and dean.

While most of the data came from the web sites of the individual iSchools, this did not provide the full data set, particularly as different schools' sites offer varying levels of detail about their faculty's credentials. Additional data needed for this analysis included the department or school granting the faculty members' terminal degree. These data were mined from the Proquest UMI Dissertation Abstracts database, faculty web pages, and faculty vitae to complete the full data set. Complete data were retrieved for all but three of 769 faculty members. The increase in population size from 674 in 2007 to 769 in 2009 is primarily due to the addition of two new iSchools at Carnegie Mellon and Singapore.

4.1.2 Classification

The areas of faculty specialization were coded into broad disciplinary areas, shown in Table 1. These categories are composed based on logical groupings of related fields of study, modified from the Classification of Instructional Programs (CIP) [8]. For example, Computing contains Computer Science along

with Electrical Engineering and Mathematics; Electrical Engineering and (applied) Mathematics are precursors of research in Computer Science, and many iSchool faculty with these degrees are trained in various aspects of Computing. The distinction between "information" and "library" studies was less clear; Communication Information and Library Studies was considered an Information degree due to the ambiguity stemming from the diversity of fields in the degree name, but all other instances where "library" occurred in the name were considered Library degrees. While scholars of almost entirely different backgrounds may receive degrees with the same name, it is impossible to distinguish in which category a given faculty member's educational experiences may better fit based on the degree names alone, so for these faculty members, we chose the more conservative classification of Library.

Table 1 Classification of Disciplinary Areas for 2009 iCaucus

Area	N (%)	Component Areas
Computing	233 (30%)	Computer Science, Electrical Engineering, Mathematics
Information	88 (11%)	Information Science, Information Studies, Information Transfer, Communication Information and Library Studies
Library	79 (10%)	Library Science, Information and Library Science, Library and Information Science
Social & Behavioral	78 (10%)	Psychology, Sociology, Social Sciences
Management & Policy	70 (9%)	Business, Management, Policy, Economics
Science & Engineering	69 (9%)	Life Sciences, Physical Sciences, Statistics, Engineering (not electrical)
Education	58 (8%)	Education
Humanities	54 (7%)	History, Philosophy, Literature, Multi & Interdisciplinary Studies
Communication	40 (5%)	Communication

Social & Behavioral disciplines included psychology, sociology and social sciences. Economics was grouped with law, business and management in the area of Management & Policy because the methods and applications of economics research in many iSchools is (arguably) more congruent with policy and strategy applications than the behavioral and social sciences. In the area of Science & Engineering, physical and life sciences are well represented, while statistics and engineering are related areas that appeared less frequently. The Humanities are dominated by historians, a number of whom specialize in science and technology studies, as well as scholars of literature, who are most common at iSchools with long-standing library programs.

5. FINDINGS

The breakdown for the disciplinary makeup of the full iSchools community in 2008, shown in Table 1, demonstrates that at the community level, the strongest area of emphasis is computing. The total number of computing-trained faculty is equivalent to the next three leading areas together: information sciences, library, and social & behavioral sciences.

The data presented in Table 2 helps make clear that the small number of schools and large variations in the compositions of the iSchool faculty across schools must be taken into account. For example, the dominance of computing in the overall picture is attributable to large numbers of computer science faculty from Georgia Tech and UC Irvine, two of the largest units in the iSchools Caucus. Likewise, the strong representation of communication is largely due to the presence of Rutgers, and in the future will be influenced by the recent merger of Florida State's iSchool and Communication department. Similarly, UCLA is responsible for the prominence of education. Computing aside, there is a fairly even distribution of scholars in 5 additional "core" areas for iSchools: management & policy, information, library, science & engineering, and social & behavioral studies.

6. DISCUSSION

The individual iSchools' histories and development trajectories have yielded a diverse set of intellectual roots that provide breadth and richness to these interdisciplinary research environments. In this section we discuss these roots, and somewhat more speculatively, how we expect they may affect the future development of the iSchools community.

6.1 Intellectual Heritage

Building from these findings, and particularly the examples of UCLA and Rutgers, data suggest that the processes of organizational emergence are one source of the community's intellectual breadth. For iSchools that undergo mergers or enter into institutional partnerships, the prior identities of the disciplinary consorts remain at least partially intact. This sort of outcome may be a result of culture, strategy, physical locations, accreditations, or even more likely, a combination of these factors – an institutional arrangements perspective.

Similarly, examining the iSchools based upon the areas of greatest concentration in faculty expertise shows evidence of the influence of "local logics" on their development. That is, the form and shape of the iSchool has more to do with response to the local situation of the school than a strong and shared intellectual identity across iSchools. One such example is Syracuse, where the strategic decision of a former Dean to establish a degree program in Information Management, in combination with a research focus on information policy, has yielded a faculty with one third of its members hailing from a management or policy background. Notably, however, there is greater sub-disciplinary diversity within the category of management and policy at Syracuse than elsewhere. Other examples of faculty compositions suggest similar local logics guiding their development, based on their unique outcomes. This local logics argument helps explain why Toronto, where the humanities are best represented as a proportion of faculty, and Penn State, where science and engineering fields are notably well represented – even with the inclusion of electrical engineering under the category of Computing – are both iSchools, but differ substantially in their faculty composition.

Table 2: iSchools' intellectual demographics in 2009.

Area of study	Total N	Overall %	Berkeley	Carnegie Mellon	Drexel	Florida State	Georgia Tech	Indiana Info	Indiana SLIS	Pittsburgh	Penn State	Rutgers	Singapore	Syracuse	UC Irvine	UCLA	U Illinois	U Maryland	U Michigan	U North Carolina	U Texas Austin	U Toronto	Washington
Computing	*233*	30%	39%	10%	27%	8%	79%	59%	9%	28%	16%	4%	70%	3%	75%	2%	7%	11%	24%	12%	9%	16%	16%
Information	*88*	11%			19%	12%	1%	3%	17%	24%	11%	19%		22%	2%	2%	27%	39%	11%	28%	18%	28%	23%
Library	*79*	10%	11%		12%	27%		2%	22%	10%		4%		9%		8%	30%	11%	11%	48%	36%	16%	29%
Social & Behavioral	*78*	10%	22%	17%	12%	8%	1%	5%	22%	10%	16%	17%		16%	6%	19%	13%	11%	16%		5%		7%
Management & Policy	*70*	9%	17%	61%	8%	12%					21%		20%	34%		2%		6%	21%		5%		10%
Science & Engineering	*69*	9%	6%	2%	8%	8%	12%	21%		21%	24%	6%	10%	3%	18%	2%			3%	8%		4%	7%
Education	*58*	8%		2%	4%	8%	4%	2%	13%	3%	5%	4%		6%		51%		11%	3%	4%		4%	3%
Humanities	*54*	7%	6%	7%	8%	12%	4%	7%	17%	3%	3%	4%				10%	20%	11%	11%		18%	24%	3%
Communication	*40*	5%			4%	23%		2%			5%	41%		6%		6%	3%		3%		9%	8%	3%
Total	*769*	*100%*	*18*	*41*	*26*	*26*	*84*	*61*	*23*	*29*	*38*	*48*	*29*	*32*	*67*	*67*	*30*	*18*	*38*	*25*	*22*	*25*	*31*

This intellectual diversity, both within and between iSchools, is undoubtedly a result of many intertwined factors playing out over time. The current faculty composition is the accumulation of these events as manifest in hiring decisions that represent a dynamic combination of organizational history, current identity, and future ambitions, to which we now turn our attention.

6.2 Intellectual Agenda

The implications for the future of the iSchools suggested by this descriptive analysis are uncertain. That noted, in general, we are optimistic for the future growth of the community. In taking this position, we do not rely solely on our faith in the greater community, but also the more objective indicators that form the academic "bottom line." There are many signs of health, such as continued faculty hiring (or very brief postponements of faculty searches) during an economic recession, and burgeoning enrollments concurrent with increasing distress over declining enrollments in the adjacent fields of Computer Science and Information Systems.

We anticipate one direction for future growth in the iSchools is through further mergers, particularly with departments of communications and mass media, as the second such partnership has emerged at Florida State since the data were collected for this study. In addition, we note the official name change at Rutgers has shifted their identity away from the explicit inclusion of library studies to the implicit inclusion of these intellectual traditions under the more flexible, though ambiguous, label of information. Although stalwart librarians and researchers with library science backgrounds may take umbrage at this identity shift, it is a reflection of changes that have been underway for years already, both within the iSchool community and more broadly.

While this study does not focus on longitudinal changes, we note one observation from the accumulated data: the proportion of faculty with degrees in the Library area is diminishing while the percentage of faculty with degree in Information is increasing, generally in the same or greater proportions. We suggest that this empirical trend has to do with the dynamic nature of the environment in which the iSchools operate; the names of degree programs are changing, and new junior faculty hires are more likely to have earned diplomas emblazoned with the word "information" instead of "library" than they were even five years ago. The iSchools are not the only institutions choosing to embrace an information-centered identity.

However, we also note that there is no evidence to support the notion that Library-focused scholarship is being phased out in the iSchool community. As previously mentioned, an Information degree may be wholly focused on library science, just as easily as it may be entirely centered on human-computer interaction or information policy. Therefore, it is entirely possible that as faculty members with Library degrees retire, their successors are also scholars of library science. Despite the shift in labeling, a rose is a rose, as the saying goes.

Another potential direction for growth in the iSchool community is through partnership, merger, or simple expansion into the field of Information Systems. These and the more narrowly defined management information systems, or MIS, departments are also interdisciplinary environments focused on applied research on computing in organizational settings, and most often found in business schools. Their research goals and scholarly interests are compatible with the work of iSchools [11] and several iSchools already include faculty from this research community.

Finally, we consider the future of the newly-minted PhDs graduating from the iSchools. While we are not aware of any

reported findings regarding placement trends for graduates of iSchools, Olson and Grudin [23] note that they are faring well in the job market, although not all are choosing to pursue employment in academia. It is not clear whether this is due to a surfeit or deficit of options for graduates; broadly speaking, the last two years' job markets have not provided as many opportunities as would ordinarily be expected under better economic conditions.

In addition, following our earlier point that interdisciplinarity requires disciplines, it is an open question as to whether the iSchool graduates themselves, as interdisciplinary scholars, are adequately grounded in disciplinary roots to become desirable faculty candidates to the iSchool community itself. Each iSchool is sufficiently distinct at this juncture that for any one school, hiring from the graduates of other iSchools is unlikely to yield a net decrease in actual intellectual diversity.

However, it also seems that, given enough time, particularistic hiring practices which unduly favor graduates from within the community would lead to greater convergence not only in the faculty's degree names, but also in the actual content of their interdisciplinary heritage, leading toward institutionalization and disciplinarity. Like so many other fields, the iSchools are likely to produce more graduates than academic jobs, so the current diversity of graduate placement seems a fair indicator of potential future trends. Some end up in policy, administration, or private sector research settings, and some remain in academia.

While the idea of homogenization of the iSchools brings mixed reactions, we note that the hiring trends to date, although providing only a brief history, suggest that this is an unlikely outcome for the near future. It seems far more likely that the iSchools will continue to focus faculty recruitment on attracting the most suitable candidates for their needs based on the institutional structures of their local environment and the particular interests and needs of their unit. This implies that there will be hiring along disciplinary lines to support programmatic needs such as professional accreditation (e.g., ALA or ABET), or selecting candidates from other iSchools who bring unique blends of expertise that complement the existing faculty research portfolio.

6.3 Limitations & Future Work

The limitations of this work include the short time scale for observing change, and the use of secondary data sources that have some known issues. Future work will include further analysis of the data used in this study, as well as ongoing monitoring of the community composition for a more useful longitudinal analysis. Incorporation of these quantitatively focused analyses with qualitative and explicitly historical accounts of the iSchools' emergence (e.g. [23]) could also provide a more complete picture of the community's development.

7. CONCLUSION

The iSchools are still in an early phase of establishing identity as a community. Although it currently represents a relatively small intellectual population, some patterns are emerging with respect to interdisciplinary community development. Computing clearly plays a large role in the community as a whole, but diversity is important as well, and there are many vibrant areas of intellectual activity in the iSchools. The richness and diversity of these broad disciplinary domains make an important contribution to the community, and the variations we observe between different iSchools' intellectual composition seem to be clearly related to local logics that, over time, have guided hiring to meet local needs. From this, we infer that these local arrangements are more important to hiring decisions than any sense of shared community identity, which is consistent with the findings of prior research on the emergence of interdisciplinary academic endeavors. These early findings reflect only a brief history of community development; however, the outlook at this time suggests that the iSchools will likely find valuable sources of fresh perspectives by pursuing new intellectual areas for growth, while continuing to cherish the important contributions of the traditional domains upon which they are building their successes.

8. REFERENCES

[1] Annabi, H., Fisher, K. and Mai, J.-E. 2005. Our Academic Life: Challenges Facing i-Schools. iConference 2005.

[2] Bair, J. 2003. Hiring Practices in Finance Education. Linkages Among Top-Ranked Graduate Programs. Am. J. Econ. Sociol. 62, 2, 429-433.

[3] Baldi, S. 1995. Prestige Determinants of First Academic Job for New Sociology Ph.D.s 1985-1992. Sociol. Quart. 36, 4, 777-789.

[4] Bedeian, A. and Feild, H. 1980. Academic Stratification in Graduate Management Programs: Departmental Prestige and Faculty Hiring Patterns. J. Manage. 6, 2, 99-115.

[5] Bruce, H., Richardson, D. J., and Eisenberg, M. 2006. The i-Conference: Gathering of the clans of information. Bulletin of the ASIST, April/May 2006.

[6] Burris, V. 2004. The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks. Am. Sociol. Rev. 69, 2, 229-237.

[7] Chen, C. 2008. Thematic Maps of 19 iSchools. In Annual meeting of ASIST.

[8] Classification of Instructional Programs (CIP 2000). National Center for Education Statistics, http://nces.ed.gov/pubs2002/cip2000/, retrieved 7/29/09.

[9] Constable, R. L. and Richardson, D. J. 2009. CRA-Deans Committee Formed. Computing Research News. 21, 3.

[10] Dillon, A. and Rice-Lively, M. L. 2006. Passing the taxi-driver test. Bulletin of the ASIST, April/May 2006.

[11] Ellis, D., Allen, D. and Wilson, T. 1999. Information science and information systems: Conjunct subjects disjunct disciplines. J. Am. Soc. Info. Sci. Tech. 50, 12, 1095-1107.

[12] Harmon, G. 2006. The first i-Conference of the i-School communities. Bulletin of the ASIST, April/May 2006.

[13] Harmon, G. and Debons, A. 2006. The i-Conference in retrospect. Bulletin of the ASIST, April/May 2006.

[14] Hildreth, C. R. and Koenig, M. 2002. Organizational realignment of LIS programs in academia: from independent standalone units to incorporated programs. J. Ed. Lib. Info. Sci. 43, 2, 126-133.

[15] Hunt, J. and Blair, J. 1987. Content, Process and the Matthew Effect Among Management Academics. J. Manage. 13, 2, 191-210.

[16] Karlqvist, A. 1999. Going beyond disciplines: The meaning of interdisciplinarity. Policy Sciences, 32, 379-383.

[17] Klein, J. T. 1985. The Evolution of a Body of Knowledge: Interdisciplinary Problem-Focused Research. Knowledge: Creation, Diffusion, Utilization. 7, 2, 117-142.

[18] King, J. L. 2006. Identity in the i-School movement. Bulletin of the ASIST, April/May 2006.

[19] Leazer, G. 2005. Split Down the Middle, Fuzzy at the Edges: Defining a Field Epiphenomenally. iConference 2005.

[20] Long, J., Allison, P. and McGinnis, R. 1979. Entrance into the Academic Career. Am. Sociol. Rev. 44, 5, 816-830.

[21] Long, J. 1978. Productivity and Academic Position in the Scientific Career. Am. Sociol. Rev. 43, 6, 889-908.

[22] Long, J. and McGinnis, R. 1981. Organizational Context and Scientific Productivity. Am. Sociol. Rev. 46, 4, 422-442.

[23] Olson, G. and Grudin, J. 2009. The Information School Phenomenon. Interactions. 16, 2, 15-19.

[24] Small, M. 1999. Departmental Conditions and the Emergence of New Disciplines: Two Cases in the Legitimation of African-American Studies. Theor. Soc. 28, 5, 659-607.

[25] Thomas, J., von Dran, R., and Sawyer, S. 2006. The i-Conference and the transformation ahead. Bulletin of the ASIST, April/May 2006.

[26] Wiggins, A. 2007. Exploring Peer Prestige in Academic Hiring Networks. Unpublished master's thesis, University of Michigan School of Information.

[27] Wiggins, A. 2009. Interdisciplinary Diversity in the iSchool Community. iConference 2009.

[28] Wiggins, A., McQuaid, M. J., and Adamic, L. A. 2008. Community identity: Academic hiring and peer prestige in the iSchools. iConference 2008.

The informatics moment: Grassrooting the space of flows in an urban branch library

By Kate Williams, University of Illinois at Urbana Champaign (katewill@illinois.edu)

This research has been funded by the Institute for Museum and Library Services and very helpfully facilitated by the Chicago Public Library.

1. Introduction

This paper examines the process whereby people seek and get guidance in using the computers and the internet in a public computer center. What is this informatics moment, as we call it, and what contributes to its success? We examine and compare the influence of both technological and social factors.

A transformation is taking place in the public library as a result of the public access internet-ready computers that patrons use. It is certainly not the first library transformation; a useful comparison can be drawn between the two concepts of the reference interview and the informatics moment. The reference interview was a 1960s reconceptualization of what takes place when a patron asks a question of a librarian, usually in order to access information in hard copy. The informatics moment as we define it is a phenomenon that entered the library along with public access computers, generally in the 1990s and since. Theorizing about the reference interview took place alongside dramatic changes in public library practice, as the popular response to the urban crisis drove people into libraries seeking help with basic needs. Then the 1980s—as the reference interview concept permeated research and teaching and practice—saw unemployment levels so high that librarians developed new library services in the form of job centers.

The dramatic shift in libraries that accompanies the informatics moment has two aspects. First, since the 1990s, libraries, particularly urban libraries, have become the last institution standing in economically decimated areas, and so they are refuges for homeless people. Second, in the 2000s libraries are central to the job search as the nation sees record numbers of jobless.

By informatics moment, we mean the interaction between a library worker and a patron as the patron seeks and gets help using the computer or internet. This is a microprocess within the information revolution where we are seeing the emergence of a new tasks and work skills and, in our data, a new (para)professional – in fact a new division of labor in the library.

Figure 1 expresses this development spatially. Where before the patron would come to the reference desk for help—the site of the reference interview—today patrons head directly for the banks of computers, or for the new worker that we found in the library. In The Chicago Public Library named this person the cybernavigator.

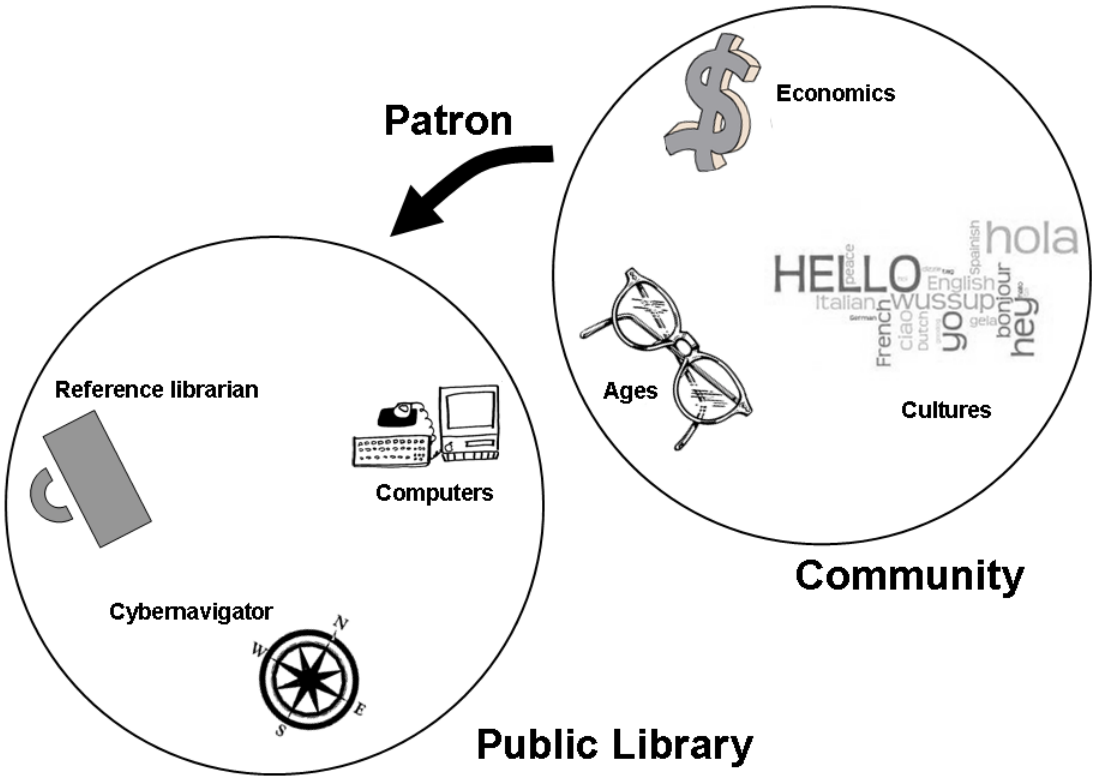


Figure 1. Library patrons reflect the economics, cultures and ages of people in a given .local community. As they enter Chicago Public Library branches, they quite often do not consult the reference librarian, but rather move directly to the computers or the nearby cybernavigator.

Libraries are important as archetypical public computer centers, where people expect service, help, education, and where the social structure of the space provides that help and the tools which make the help possible. In this study, success in the informatics moment is the extent to which people can get help with a variety of computer/internet uses. This indicates the extent to which people are helped across some particular digital divide that is an obstacle for them.

2. Literature review

The digital divide—people's differences in their access and use of information technology—has attracted a great deal of money, effort, and criticism. Research has established that the digital divide follows the demographic pattern of earlier socioeconomic divides of income, education, age, and gender (NTIA 1995, 1998, 2000, 2001, 2003, 2007), and that it is not a single divide but multidimensional and ever-changing (Clement and Shade 2001, Dimaggio and Hargittai 2001). Dimaggio and Hargittai highlight social support as one dimension of digital inequality. Van Dijk (2005) distinguishes four aspect to the digital divide, three of which are social rather than technological:

> 1) Lack of any digital experience caused by lack of interest, computer fear and unattractiveness of the new technology ('psychological access');
> 2) No possession of computers and network connections ('material access');
> 3) Lack of digital skills caused by insufficient user-friendliness and inadequate education or social support ('skills access');
> 4) Lack of significant usage opportunities ('usage access').

In theorizing the emergence of the network society, Castells (1999) has posited the space of flows as a mechanism for controlling the space of place, and has proposed grassrooting the space of flows as that process whereby non-elites join elites in online and related spaces and activities, in a cacophonous competition for shaping and controlling the world that we all live in. For close to two decades, community informatics research has examined and at times in turn influenced the unfolding of that grassrooting process. Williams and Durrance (2009) describe this process as continuity (local historical communities) encountering transformation (information technology); this definition conveys its disruptive nature.

One of the spaces of place that research has examined is the public library, with surveys at a macro level (for instance, Williams 2000) as well as with ethnographies at the micro level (for instance, Sandvig 2006). Bertot and McClure and their team offer (most recently in Bertot, Jaeger, McClure, Wright, and Jensen 2009) more than 15 years of examination of the public library as a public computer center, primarily but not only via survey data. Their work documents, primarily from surveys of library equipment, but also library processes, that library patrons are entering the space of flows, many for the first time, in an effort to control their own lives and fortunes. This process is happening in schools, at work, at home, even in the living rooms and basements of friends who have the digital tools.

This paper focuses on the microprocess of the informatics moment within the public library. What is the process? What maximizes the process? What explains the variation in the types of informatics moments? Are these moments driven by technology, or by social factors? Our conceptual framework for those social factors is social capital, that is, resources available across social networks, as in Lin (2001). This framework has been fruitful in a number of community informatics studies, reviewed in Williams and Durrance 2008).

3. Method

The setting for the research is the library system of the City of Chicago, which in 1999 begun to create a particular corps of workers who provide the help and guidance that enables people to use the library's computers and internet. According to Bertot et al (2009), nearly 100% of public libraries provide public access computers and internet. While many rural locales are outside of any service area, public libraries remain one of the largest and most robust channels of provision of public computer access.

The personnel we found work as cybernavigators in Chicago Public Library, although they are paid by the library's foundation. As of 2009 there are 40 cybernavigators across the 79 sites of the library system, particularly where demographics are associated with the digital divide, that is, lower rates of ownership and use of computers and the internet. CPL has 77 branches, two regional libraries, and a man library. The cybernavigators are each assigned to one of 37 branches or the regional or main libraries.

The cybernavigator program consists of people working 20 hours a week as independent contractors. Their task is to help people use computers and the internet, which includes teaching classes, working one-on-one with patrons by appointment and walk-in, answering questions and performing basic troubleshooting on the computers and printers. They often also help staff in the branches with computing and other needs.

Different library systems have different approaches to staffing and providing this service. Ann Arbor District Library, for instance, at one time assigned this task to their pages. The cybernavigators we encountered are not masters-degreed librarians (with one exception) and do not have reference training.

The field work is just concluding and this report is preliminary. Thus far, we have:

- examined CPL's archives and working files on this program, including reports from all 40 current cybernavigators
- observed six cybernavigators at work in different libraries
- carried out focus group discussions with 27 cybernavigators
- collected answers to questionnaires from those 27 cybernavigators

The sample of six sites was selected in order to emphasize lower-income branches and a range of ethnicities; the sample of 27 included those of the 40 who were able to make time for the researchers.

The analysis now underway combines trend data with narratives to build up a thick description that holds true across the 27 cybernavigators, or rather the 27 settings for the informatics moment.

4. Quantitative trends

We explored two models as follows:

(1) The technological facility of cybernavigator is associated with a wider range of informatics moments, or CN IT → IM.

(2) Increased social capital supporting the patron is associated with a wider range of informatics moments, or SC → IM.

To test these models, we used cybernavigators' answers to closed end questions regarding three aspects of the process. First, the help they provide to people. Second, their own technology uses. Third, the social relationships between the patron, the library, and the cybernavigator. Quantitative measures of the variables in the two models were constructed as follows:

- IM is a count of how many types of informatics moments each cybernavigator reports as going on daily in his or her branch.

- CN IT is a count of the types of technologies a cybernavigator reports using in his or her life, at work or beyond. The lowest count was 13, the highest 23, with a mean of 19; this was from a possible total of 27 yes/no questions.

- SC is a count of indicators of social capital as reported by each cybernavigator. Some indicators reflect library social capital; others reflect the cybernavigator's own social capital. Library social capital is a count of the indicators of social capital relating to the library staff that a cybernavigator experiences daily. Cybernavigator social capital is a count of the indicators of social capital that a cybernavigator reports experiencing daily.

The CN IT and IM variables show a correlation of .22 (n = 21). This suggests a minor direct relationship between a cybernavigators technology uses and the breadth of informatics moments that occur daily in his or her library.

The Library SC and IM variables show a correlation of .38 (n = 27). The Cybernavigator SC and IM variables show a correlation of .37 (n = 27). This suggests a stronger direct relationship between social capital in evidence as familiarity, trust, and helping behaviors

among community, library staff, and cybernavigator on the one hand, and the range of informatics moments that occur daily in the library.

These variables are carefully calculated from specific reported behaviors that are detailed below and at the same time the tendencies are modest. The narrative data from the focus groups, however, complements and fills in the details regarding the informatics moment, the cybernavigator's technology activities, and the social capital at work.

5. The informatics moment

Robert Taylor (1968) led in reconceptualizing what happens when a patron asks a question of a librarian. Through his work and others, that process or that moment became known as the reference interview. He interviewed librarians about that question-and-answer process and derived four stages of information need that became the basis of further research, and then of library training (Ross et al 2002). This work examined the process of getting to information that was generally in hard copy, asking a question of a library or other information system when you don't know that system.

This study analyzes data collected from a new type of library worker: the cybernavigator. Rather than guiding people to information in hard copy, their work focuses on helping people use computers and the internet. Patrons don't have an information need so much as an objective that involves computers and the internet and they need help using the technology. We call the interaction between the patron and the cybernavigator the informatics moment, with conscious reference to the earlier examination of the reference interview. Our sense is that an examination of this microprocess at the heart of the digital divide can be fruitful for research and for practice, in libraries and beyond. A situation where library staff are unable to support a new library service, and non-librarian, non-payroll individuals are placed in the library to do this new work, demonstrates a sort of creative destruction (Schumpeter 1942/1975) that is perhaps quite appropriate for a large, rule-bound public library system in a moment of chaotic transformation.
So library worker and patron interact, not in the reference interview, but in the informatics moment. What is this moment? The data suggests a typology that is laid out in Table 1. First, there is basic literacy. On average, cybernavigators report helping people with reading or writing skills on a weekly basis. Second, there is computer literacy, which is either basic or advanced. Third, there are functional activities that people are getting help with. Some of these functions are specific to the library and can be defined as library literacy: navigating the unfamiliar or tricky print process and the computer reservation system. Still others are functional online activities reflecting patrons' lives.

Table 1 below ranks various types of patron activities that are involved in this informatics moment according to how frequently (on average) the 27 cybernavigators experience them.

	Literacy	Computer literacy	Advanced computer literacy	Library literacy	Functional activity
Daily		using the mouse or browser, getting or using email		printing, computer reservations	searching for work, applying for jobs
Weekly	writing reading		producing or updating a document besides resume, doing research, using other government websites besides benefits, using social networking sites	using the library catalog, using library databases	producing or updating resume, doing homework, getting or checking benefits
Monthly		taking a computer class	playing a game		looking into current events or cultural information, getting health information, banking/buying/selling/ other e-commerce, seeking resources relating to being homeless

Table 1. The informatics moment: What kinds of help are patrons seeking and getting from cybernavigators?

The most common informatics moments, encountered on average daily by the 27 cybernavigators, involve either basic computer tasks such as using a mouse, browser, or email; complex library-specific tasks such as printing or making reservations, and activities such as looking for work or applying for jobs.

Informatics moments across the 27 cybernavigators vary a great deal. Three cybernavigators report only two kinds of informatics moments occurring daily. In each case these are helping people print and handle their computer reservations. At the other extreme, one cybernavigator reports 19 kinds of informatics moments happening daily. The average across all cybernavigators is 10.

As one Champaign, Illinois, computer instructor tells his beginner students, "The digital divide is something that happened while you were at work." Today, newly unemployed people encounter a job market that has migrated online. Businesses with semi-skilled and even unskilled job vacancies point jobseekers to the public library to fill out an application. Out of dozens of stories cybernavigators told about job seekers, many stand out, and here is one. (All unattributed quotes are cybernavigator comments in focus groups.)

> One of my best success stories was a young lady I was assisting with job searching. We originally started out with just the internet basics and, you know, she just wanted to learn how to use the computer. And it went from learning how to use the computer, then she wanted to do a resume, and then – she seen me helping someone else do a resume prior to her session, and then she was interested in wanting to make a resume. We went from resumes to applying for

> jobs online. And she was coming out of a battered relationship, and it went from – she wanted to have her own identity, her own income, and she went from, like, not really having a job or having a lot of experience, but I was able to utilize just her volunteer experience and some of – you know, just some of the life experiences that she had, to be able to articulate it into a resume format. And we applied for a job at ---- as an environmental specialist, and she actually got the interview and went in and got the job and everything. And I was just so, so happy for her because she was like: "This is, like, the most money I ever made in my life," and she's like: "I would never have made it if it wasn't for you," and everything, and I just—I felt good because she was looking for something and she got it, you know what I'm saying?

At a more basic level, printing and computer reservations are frequent challenges for patrons because they are complicated.

> You have to go up to the front desk, give them your card and some money. But then if you don't remember how many pages you're printing, you got to go back to the print station, click on it, type in your number, find out how much you owe, go back up to the front desk, add the money to the card, then go back over there, click on the website, which… There's nothing telling you to click here. It's just, how do people know?

Reservations are made at a dedicated station:

> It's kind of sort of – a kind of complex situation to be able to go make a reservation. If you don't have your library card number, you need to go back up to the front desk, have a clerk get your library card number or apply for a library card, go back to the reservation station, make a reservation. You might have to wait an hour or two for a reservation to get a full hour session. But then they see empty seats because you get a 10 minute grace period.

In just these three examples the social aspects of help—the cybernavigators received a day or so of training that called it customer service—loom large, within the environment of important but scarce resources (computer time, prints) that is the public library.

6. Technology expertise among cybernavigators

We operationalized the technological facility of the cybernavigators as how many of 27 IT activities they reported doing. Cybernavigators reported engaging in between 13 and 23 of these. Table 2 indicates how many engage in each activity; 21 activities are universal or common; six are relatively rare.

Table 2. Cybernavigator technology use: At work or otherwise, do you…

talk on a cellphone	100%
create documents on a computer	100%
use wireless to connect to the Internet	95%
use Wikipedia	95%
text on a cellphone	95%
send or receive e-mail as part of group activities	95%
look for information on the Web	95%
share photos, audio or video or that you have made	90%
use instant messaging	86%
use a spreadsheet	86%
use online chat	81%
take digital photos	81%
post to an electronic discussion list or bulletin board	81%
send/receive email on a cellphone	76%
read an online bulletin board	76%
post information on the Web in some other way, blogging for instance	76%
browse the web on a cellphone	76%
belong to an electronic discussion list	76%
talk over the Internet as you would on a telephone (e.g. Skype)	71%
record digital video	62%
record digital audio	52%
use bookkeeping software	33%
host or edit an electronic discussion list or bulletin board	29%
create or maintain web pages	29%
use Linux or any open-source software	24%
add to or change a Wikipedia entry	19%
write a program	14%

N = 21 cybernavigators

In addition to these closed end question answers, there is a great deal more on this topic in focus group transcripts, but two facts stand out. First, while all but one of the cybernavigators report using Wikipedia, only four are contributors to Wikipedia. Elsewhere (Williams 2005) we have noted the difference between downloaders and uploaders, or what Castells (2001) calls the interacting and the interacted. With Wikipedia one of the most consulted reference works, can library workers serve patrons well if they aren't also part of creating that new and complex tool?

Second, more than one cybernavigator described themselves with regret as just barely ahead of their patrons with respect to technology skills, without time or access to learn new skills. They took for granted their demonstrated ability to think and learn—and

teach!—on their feet. Their most important skill, many did reported with humor and pride, was patience—the ultimate people skill.

7. Social capital

Past work (Williams 2005) suggests that social capital is a determinant of IT use; in other words, greater quality and quantity of social support facilitates greater IT use.

For model (2) above, we operationalized social capital in two dimensions: cybernavigator social capital and library social capital. Table 3 details the indicators that sum up to those two dimensions and, as in table 1, how often (on average) the cybernavigators witness each.

	Cybernavigator social capital	Library social capital
Daily	CN helps someone who knows his or her name	Librarians connects patron to the CN for help Paraprofessionals connect patron to the CN for help
	CN helps someone he or she knows by name	
	CN helps someone he or she recognizes but don't know by name	
Weekly	CN helps someone that a community member referred to him or her	Security guard connects patron to the CN for help
		Non-CN library staff help someone with computer/internet, apart from printing/reservations
Monthly	CH helps a group of 2 or more working together on a task	[none]
	CN helps someone he or she knows from some community involvement	
	CN brings their own laptop to work	
	CN shares his or her own laptop with patron as part of help	
Less than monthly	CN sees, runs into, or gets together with patrons outside of work	CN sees, runs into, or gets together with library co-workers outside of work

N = 27 cybernavigators

Table 3. Frequency of various indicators of social capital.

The narratives regarding social interactions and relationships are complex and rich. Just two examples. Some branch managers frown on cybernavigators bringing in laptops, and some don't have laptops to bring in. But those who do bring them in celebrate their ability use them to learn and stay ahead of patrons, and one of the most busy cybernavigators actually turns her laptop around for up to four patrons at a time to look at

and use together. The tools on her laptop are what the patrons want to learn. Many patrons also come in for help with their own laptops.

Second and perhaps more important is the dynamic of other people helping patrons besides the cybernavigators. Library staff referring patrons to the cybernavigator is critical and valued, and it reflects trust in the cybernavigator. But it can happen without regard for whoever the cybernavigator is already helping. Or help can be offered by library staff or even by a fellow patron—and it is incorrect. This conflict and confusion as everyone tries to help reflects the creative destruction of formal roles as the library tries to meet current demands. It may be that these role changes can be consciously directed in order to move towards more effective service that relies on both paid and unpaid help—a third time of social capital, community social capital—in a time of extremely tight budgets.

Certainly, oftentimes the demands on the cybernavigator exceed those on library staff, and this is uncomfortable and leads to further conflict. But sending people to the cybernavigator increases the functionality of the library. If the creative destruction had created too much hostility, and cut off the CN, then it wouldn't have been effective. This we did not find. On the other hand, from the point of view of the cybernavigator, what has to happen is the entire library has to become more connected and able to help patrons. Cybernavigators asked for greater functionality of the technology—word processing software besides in the browser, flash drives. The capacity of the cybernavigator and the demands of the patron currently exceed the technological capacity of the library.

8. Conclusion, implications, and next steps

As stated, this is a report of preliminary results. We have demonstrated a tendency for social capital to be more influential than technological skill in the informatics moment. The precoded survey data provided quantitative evidence and the stories from the focus groups and ethnographic observations spelled out a nuanced interpretation on a qualitative level.

Our intention is to explore further the social aspects factors of the community as well as the library location itself, and to document in greater detail the experience of the cybernavigators in these informatics moments.

The significance of this is that if we can better understand the relative importance of technology and social capital, we can better design programs of social intervention to end the digital divide. Our past research (Alkalimat and Williams 2001, Williams and Takazawa 2008) indicates that people with resources in their own communities make use of them. These resources, including local public libraries, are the most useful place to start.

Key references

Alkalimat, A., and Kate Williams. Social Capital and Cyberpower in the African American Community: A Case Study of a Community Technology Center in the Dual City. In *Community Informatics: Community Development Through the Use of Information and Communications Technologies*, edited by Leigh Keeble and Brian Loader, London: Routledge, 2001.

Bertot, John; Jaeger, Paul; McClure, Charles; Wright, Carla, and Jensen, Elise. "Public libraries and the Internet 2008-2009: Issues, implications, and challenges" *First Monday* [Online], Volume 14 Number 11 (24 October 2009)

Castells, Manuel. Grassrooting the Space of Flows. *Urban Geography* 20 (4) 1999 p 294-302, reprinted in *Cities in the Telecommunications Age*, James O. Wheeler, Yuko Aoyama, and Barney Warf, eds. p 18-27.

Castells, Manuel. *The power of identity*. Malden Mass.: Blackwell, 1997.

Dijk, Professor Jan A G M van. The Deepening Divide: Inequality in the Information Society. 1st ed. Sage Publications, Inc, 2005.

Sandvig, C. (2006). The Internet at play: Child users of public Internet connections. *Journal of Computer-Mediated Communication, 11*(4), article 3. http://jcmc.indiana.edu/vol11/issue4/sandvig.html

Schumpeter, Joseph A. *Capitalism, Socialism and Democracy* (New York: Harper, 1975) [orig. pub. 1942]

Williams, Kate, and Joan C. Durrance. "Community Informatics." In *Encyclopedia of Library and Information Science*. Taylor and Francis, 2009.

A complete bibliography is available from the author and will be online for the i-schools conference.

Effective ICT Use for Social Inclusion

Martin Wolske
University of Illinois Urbana Champaign
501 E. Daniel St.
Champaign, IL 61821
+1-217-244-8094
mwolske@illinois.edu

Noelle Sheree Williams
Same address and phone
willi102@illinois.edu

Eric Orace Johnson
Same address and phone
eojohnsn@illinois.edu,

Robin Yoerger Duple
Same address and phone
duple1@illinois.edu

Safiya Umoja Noble
Same address
+1-217-819-7648
snoble@illinois.edu

ABSTRACT

Access to information and communications technology (ICT) is considered important for individuals to fully achieve educational and economic development goals. In fact, ICT access has become so important that the lack of it has been termed the digital divide. To combat the digital divide, community-based computing centers were created as a vital first step to provide physical access to ICT. Commonly known as Community Technology Centers (CTCs) or Telecentres, these publicly accessible labs are providing a valuable means for the diffusion of technology in underserved communities. The meaningful digital divide, however, is whether individuals can fully participate in society. Access to tools like computers and the Internet are only the first step toward effective social inclusion. Furthermore, if the tools become the focus, and we look exclusively at diffusion of technology to address the digital divide, then we make compromises in implementation that never address other, equally important, issues.

If public computing facilities such as CTCs are to make the transition from facilities fostering diffusion of technology to community centers empowering citizens through effective use of ICT as citizen professionals such as citizen scientists, citizen planners, and citizen journalists, it is necessary to revisit implementations of technology within these spaces; this may mean creating a new framework for how computers and other ICTs are set up for use in CTCs and Telecentres. Techniques used by African-American marketing agencies as well as successful non-profit organizations that implement grassroots campaigns can teach us a lot about designing compelling experiences that attract audiences.

General Terms

Management, Design, Experimentation, Human Factors

Keywords

"Community Technology Center (CTC)," "Community Informatics," "Citizen Professional," "Mass Amateurization," "Public Computing Space Design Aesthetics."

1. INTRODUCTION

Information and communications technologies (ICT) have become an important tool helping individuals fully achieve their educational and economic development goals. The digital divide is the inability of underserved populations to access and use ICT, furthering social, economic, and educational inequities [16, 17, 18, 19]. To combat the digital divide, community-based computing centers have been used as a vital first step to provide physical access to ICT. Commonly known as Community Technology Centers (CTCs) or Telecentres, these publicly accessible labs are providing a valuable means for the diffusion of technology in underserved communities.

The meaningful divide, however, is whether individuals can fully participate in society. Gurstein [9] points out that achieving educational and economic development goals in an information society requires more than simple physical access to ICT; access to tools like computers and the Internet are only the first step toward effective social inclusion. Furthermore, if the tools become the focus, and we look exclusively at the issue of diffusion of technology to address the digital divide, then we make compromises in implementation that never address other, equally important, issues. According to Fuchs [11], while diffusion of technology is often the initial impetus for creating CTCs, this need is reduced over time as more access is available in private residences. At that point CTCs either must transform to meet other "back of the market" needs, or dissolve. Examples of such transformations can be found in countries like South Korea where even though the percentage of residents with home computers and broadband Internet access is very high, public computing facilities such as cyber-cafes still thrive as social hubs. Further, public physical spaces can foster behaviors and activities that may lead to greater civic participation by hosting a complex range of interactions with ICT, including some activities that also take place in private computing spaces [29].

While it may be necessary to start with an emphasis on diffusion, as rapidly as possible we need to emphasize effective use for social inclusion. To make this shift, we need to consider what roles and activities technologies need to support. While the diffusion focus often results in technological implementations that help meet basic educational needs and prepare youth and adults for entry-level positions (with few options for advancement), the effective use for social inclusion model results in technology

implementations that help prepare residents to be leaders in building stronger communities and shapers of the workplace as opposed to being employed by those shaping the workplace.

The question then is, how do we implement technology in a way that ultimately prepares residents to be leaders? A framework of effective use is needed, defined by Gurstein as "[t]he capacity and opportunity to successfully integrate ICTs into the accomplishment of self or collaboratively identified goals." [9] In order to create this framework for underserved populations, we must first understand the roles and activities that lead towards a stronger civil society, the role of a citizen professional, as well as the ICTs that a community member needs to assume that role. Secondly, we must also understand the current trends in the design of public computing spaces in order to project how technology may function in the future and how ICT will affect communities. By understanding the various roles of the citizen professional, the current trends in technology, and the design of public computing spaces, we will be able to amplify community voices through ICT.

2. THE CITIZEN PROFESSIONAL

According to Vaughan [28], "[t]he priority of the scientific community and government should be empowering citizens, developing the tools and approaches to bridge the gap between civil society and decision-makers." This is exactly the idea behind the citizen professional, a community member who plays the role of an amateur scientist, urban planner, or journalist within the community and whose efforts allow the community to be a part of major decision making processes that affect the community.

In physically bringing together community members, each with different backgrounds, insights, and skills, community technology centers can serve as places where knowledge acquired by one can be shared with others, an important aspect of any sustainable CTC. These centers thus allow community members to participate in citizen professional roles and become gathering places for "communities of practice" [34], that is, groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly. CTCs can also serve to foster communities of inquiry, groups united by shared interests who work together to investigate and act to address common problems [25]. Based on the early-twentieth-century theories and practices of Charles Peirce, John Dewey, and Jane Addams, a number of researchers have proposed that the inquiry cycle, especially as applied to communities of inquiry, provides a rich environment for achieving educational and community development goals [5, 4, 6, 26, 25, 33]. Indeed, these examples illustrate the importance of creating centers that first and foremost emphasize community collaboration in a physical space while providing the necessary ICT to support such activities.

Indeed, the desire to build social capital is something many CTCs (or Telecentres) have in common. Ceballos [7] suggests that "the best Telecentres are local gathering places; places where people come together to talk, tell stories and share knowledge." As stated earlier, public and private physical spaces can foster behaviors and activities that may lead to greater civic participation. Depending on the setting, members of the community may enter into conversations with each other to determine what is needed to improve the local quality of life. Thus, not only do they become actively engaged in adapting tools to their own local needs, but also they become more connected with each other, possibly further strengthening the community. CTCs serve as both hubs where people join communities of inquiry to determine what is needed to improve their quality of life and centers where the tools necessary to implement some of those changes are available for community use. New technologies enable self-determined media production that multiplies the voices in communities. Sharing local stories may forge civic bonds. In this way, communities are better able to define their own future.

Berry, et. al. [1] points out that participation in the political process is best fostered through regular face-to-face interactions. The strongest local governments are the ones that create mechanisms to encourage the formation and bi-directional flow of information with healthy neighborhood-based associations. Such associations help balance power between the elite and non-elite within the community and are shown to decrease tensions between those advocating for neighborhood or business interests. As local information hubs and gathering places, CTCs clearly have the potential to move beyond serving as merely a channel for the delivery of government services to being a platform that helps provide communities with skills needed for a new type of citizenship. They become places to negotiate the future of governments creatively and inventively [7]. This new type of citizenship and the negotiation of future government are facilitated by the presence of citizen professionals in the community.

We see examples of this trend in high-functioning communities. They are finding ways to take advantage of the mass amateurization brought about by emerging technologies, such as low-cost ultra mobile personal computers, smartphones, cloud computing, geographic/neighborhood information systems, and personal webs, to engage community members in community development goals. According to the 2009 Horizon Report, newly emerged and emerging technologies such as mobile and cloud computing are leading to a collective intelligence and mass amateurization that are redefining how we think about virtual and physical spaces. Cloud computing, networked computers that distribute processing power, applications, and large systems among many machines [10], provide a unifying technological base for grassroots video and collaborative webs that are empowering citizen professionals at many levels. The Horizon Report also reviews a number of ways in which the growing availability of Geographic Information Systems (GIS) targeted at the consumer market is being used by citizens to enhance storytelling, health, and learning objectives. Already considered as another component of the network on many campuses, mobile devices continue to evolve rapidly. New interfaces, the ability to run third-party applications, and location-awareness have all come to the mobile device in the past year, making it an ever more versatile tool that can be easily adapted to a host of tasks for learning, productivity, and social networking [10]. These technologies enhance the overall quality of outcomes by using a diversity of input, but equally important to these virtual collaborations are the physical spaces of CTCs where some collaborations take place. The challenge remains to empower more communities, especially in underserved areas, through access to, and training with, citizen professional ICTs, as well as through a design of space that is welcoming and engaging to draw in community interest and participation.

A training or educational component of promoting citizen professional ICTs is also very important to meet the needs of the clientele of the CTC or Telecentre, some of whom may have had little previous exposure to technology. Communities of practice may share information and help each other learn to perform tasks better through interaction with others in the community, but this assumes a basic level of skills that some or all members possess in order to be able to teach them to one another. Fuchs [11] points out that the best way to do technology training in CTCs for less experienced and first-time users is often through "individual training or small group sessions." But where do the people who have these skills to be able to pass along come from? In many cases, the Telecentre or CTC staff members are the first in a community to possess the technological skills needed. According to Fuchs, Telecentres create a capacity for technological prowess that "becomes self-generating" [11] as staff pass along their skills and empower community members. These community members may then pass their newfound skills on to still more community members through interaction and collaboration, often becoming innovators and early adopters as a result of the skills and confidence learned in the Telecentre.

3. DESIGN OF LOCAL INFORMATION HUBS AND GATHERING PLACES

The design of spaces in low-income neighborhoods is understudied but important for understanding the psychographic traits and needs of underserved communities. The ways in which spaces regulate social experiences and constrain individual and social relations is an important topic in the development of public computing centers. CTCs must support the ways ICTs will be used to encourage citizen professional activities. While numerous professionals find mobile devices the platforms of choice for their daily work lives, many CTCs are still built in closed rooms and restricted to mass implementation of non-flexible technology directed only at basic bridging of the digital divide.

If public computing facilities such as CTCs are to make the transition from facilities fostering diffusion of technology to community centers empowering citizens through effective use of ICT as citizen professionals such as citizen scientists, citizen planners, and citizen journalists, it is necessary to revisit implementations of technology within these spaces; this may mean creating a new framework for how computers and other ICTs are set up for use in CTCs and Telecentres. Techniques used by African-American marketing agencies as well as successful non-profit organizations that implement grassroots campaigns can teach us a lot about designing compelling experiences that attract audiences. Design matters because it directly impacts our real and perceived quality of life and experience, and the design of public computing spaces will be an important contribution to the implementation and approach of community informatics.

Design and aesthetics that resonate with the aspirations of African-Americans, for example, are of critical importance when cultivating ongoing relationships. Companies are increasingly using experiential marketing companies to cultivate emotional relationships with brands and products that rely heavily upon beautiful spaces and intriguing environments. Social spaces meet a variety of instrumental and emotional needs, and the public computing center can be evaluated with this in mind. People use public computing spaces to check email, apply for a job or work on a computer to accomplish a variety of tasks. These are the instrumental aspects of this public space. But there are also the non-instrumental or emotional aspects of the social experience in public computing spaces. Human-computer interaction literature focuses on the experience of users with technology but neglects the study of the environment in which human-computer interactions take place. Interface design happens at the level of individual interaction with the machine. But spatial design occurs at the individual and social level. Spatial design impacts the types of interactions that are allowable, both with machines and with other users in the public computing center. The intersections between current designs of public computing space with what we can learn from experiential marketing efforts targeted at African-Americans provide some insights into the crucial intersections of virtual and physical spaces.

4. CONCLUSION

CTCs are often seen as bridges between basic ICT services and the superior but more expensive access to ICTs in places of residence. However, there is a range of examples that exist in which CTCs are transforming. They are being used not only by those who do not have access elsewhere, but also by those who have access but value the social interactions found at these community centers. In short, they are shifting from places of technology diffusion to places that promote social inclusion, and encourage communities of practice and inquiry to form. High functioning neighborhoods and organizations find ways to bring together citizen professionals in communities of inquiry/practice to help build civil society. Today, this also means finding ways to equip people with, and inform them about, emerging technologies that support mass amateurization.

Such trends show us that as we build CTCs for underserved areas, we must create them in a way that encourages similar citizen professional activities if we are to truly foster social inclusion. We must also examine trends in the ICTs currently used by various types of citizen professionals, communities of practice and communities of inquiry, and study the impact of experiential marketing targeted to urban consumers and community members to improve both the public and social computing experience in CTCs. Only then can ICTs that are most appealing to these community members in CTC settings be combined with the design of physical spaces in ways that will encourage residents to take an active role in shaping and leading their own communities.

5. REFERENCES

[1] AGARWAL, A., AND MEYER, A. 2009. Beyond usability: evaluating emotional response as an integral part of the user experience. In Proceedings of the 27th International Conference on Human Factors in Computing Systems, Boston, MA, Apr. 2009, ACM Press, New York, NY, 2919-2930.

[2] ALLEN, M.W. 2008. A dual-process model of the influence of human values on consumer choice. A revista Psicologia: Organizações e Trabalho (rPOT) 6, 1, 15-49.

[3] BERRY, J. M., K. E. PORTNEY, AND K. THOMSON 1993. The Rebirth of Urban Democracy. Brookings Institution, Washington DC.

[4] BISHOP, A., AND BRUCE, B.C. 2008. Liberating Voices! A Pattern Language for Communication Revolution: Community Inquiry [Online]. Available: http://www.publicsphereproject.org/patterns/print-pattern.php?begin=122.

[5] BRUCE, B.C., AND BISHOP, A.P. 2008. New literacies and community inquiry. In Handbook of Research on New Literacies, J. COIRO, M. KNOBEL, C. LANKSHEAR AND D.J. LEU, Eds. Routledge, New York, NY.

[6] BRUCE, B.C., AND BISHOP, A.P. 2002. Using the Web to support inquiry-based literacy development. Journal of Adolescent & Adult Literacy 45, 8, 706.

[7] CEBALLOS, F. ET AL. 2006. From the ground up: the evolution of the telecentre movement. Ottawa, ON, CA: Telecenter.org / IDRC, accessed July 2009 at http://idl-bnc.idrc.ca/dspace/handle/123456789/27550

[8] COCKTON, G. 2008. Designing worth - connecting preferred means to desired ends. Interactions 15, 4, 54-57.

[9] GURSTEIN, M. 2003. Effective Use: A community informatics strategy beyond the Digital Divide. First Monday, 8, 12, 1 December 2003.

[10] JOHNSON, L., LEVINE, A., AND SMITH, R. 2009. The 2009 Horizon Report. The New Media Consortium, Austin, Texas.

[11] FUCHS, R.P. 1998. Little Engines That Did--Case Histories from the Global Telecentre Movement. Ottawa, ON, CA: Telecenter.org/IDRC, http://www.idrc.ca/fr/ev-10630-201-1-DO_TOPIC.html.

[12] HASSENZAHL, M. 2006. User experience (UX): towards an experiential perspective on product quality. In Proceedings of the 20th French-speaking Conference on Human Computer Interaction (Conférence Francophone sur l'Interaction Homme-Machine) IHM '08. Metz, France, September 2008, ACM Press, New York, NY, 11-15.

[13] HASSENZAHL, M., AND TRACTINSKY, N. 2006. User experience – a research agenda. Behaviour and Information Technology 25, 2, 91-97.

[14] MITTAL, B. 1988. The role of affective choice mode in the consumer purchase of expressive products. Journal of Economic Psychology 9, 499-524.

[15] NAFUS, D. 2009. From multitasking to plastic time: the busyness (or not) of technology use. Seminar presented January 23, 2009, Abstract accessed July 2009 at https://apps.lis.illinois.edu/wiki/display/sp09lis590ul/Schedule

[16] National Telecommunications and Information Administration (NTIA), US Department of Commerce 2002. A nation online: how Americans are expanding their use of the internet. Washington, DC: NTIA.

[17] ———. (2000a) Falling through the net, toward digital inclusion. Washington, DC: NTIA.

[18] ———. (2000b) National telecommunications and information administration annual report.

[19] ———. (1999) Falling through the net: Defining the digital divide. Washington, DC: NTIA.

[20] NORMAN, D. A. 2002. The design of everyday things. MIT Press, New York, NY.

[21] PHOTOVOICE, Cameroon, November 2004 - January 2005, Photography by disabled people in the UK, Bangladesh and Cameroon, [Online] Available: http://www.photovoice.org/html/exhibitionsandevents/upcoming/bangladeshexhib.html.

[22] RAINIE, L., AND ANDERSON, J. (2008). The future of the internet III.[Online].Available: http://www.pewinternet.org/Reports/2008/The-Future-of-the-Internet-III.aspx

[23] SADUN, E. (2009). Consumers, not providers, ready for ubiquitous cellular data.[Online]. Available: http://arstechnica.com/telecom/news/2009/03/going-ubiquitous-with-cellular data.ars

[24] SAWHNEY, N (2009). Voices beyond walls: the role of digital storytelling for empowering marginalized youth in refugee camps. In Proceedings of the 8th International Conference on Interaction Design and Children. New York: ACM Press, 302-305

[25] SHIELDS, P. (1999, Mar.). The Community of Inquiry: Insights for Public Administration from Jane Addams, John Dewey and Charles S. Peirce. Public Administration Theory Network. [Online]. Available: http://ecommons.txstate.edu/polsfacp/3.

[26] SHORT, K.G., SCHROEDER, J., LAIRD, J., KAUFFMAN, G., FERGUSON, M.J., AND CRAWFORD, K.M. 1996. Learning together through inquiry: from Columbus to integrated curriculum, Stenhouse, Portland, ME.

[27] STOECKER, Randy (2005). Research methods for community change: a project-based approach. Sage Publications, Thousand Oaks, CA.

[28] VAUGHAN, HAGUE 2007. Citizen science as a catalyst in bridging the gap between science and decision-makers. In the Proceedings of the Citizen Science Toolkit Conference, Ithaca, NY, Jun. 2007, Cornell Lab of Ornithology, Ithaca, NY.

[29] VISEU, A., CLEMENT, A., ASPINALL, J., AND KENNEDY, T.M. 2006. The interplay of public and private spaces in Internet access. Information, Communication & Society 9, 5, 633–656

[30] WANG, C., AND BURRIS, G., M.A. 1997. Photovoice: concept, methodology, and use for participatory needs. Health Education & Behavior, 24, 3, 369-387.

[31] WANG, C., ET. AL. (1998). Photovoice as a participator health promotion strategy. [Online]. Available: http://heapro.oxfordjournals.org/cgi/reprint/13/1/75.pdf

[32] WEISER, M. 1991. The computer for the 21st century. Scientific American 265, 3, 94 –104

[33] WELLS, G. 2001. Dialogic Inquiry. Cambridge University Press, New York, NY.

[34] WENGER, E. C., AND SNYDER, W. M. 2000. Communities of practice: the organizational frontier. Harvard Business Review, Jan./Feb.,139-145.

[35] WILDERMAN, C. (2007). Models of community science: design lessons from the field. [Online]. Available: http://www.birds.cornell.edu/citscitoolkit/conference/proceeding-pdfs/Wilderman%202007%20CS%20Conference.pdf

The Usages and Expectations of Multilingual Information Access in Chinese Academic Digital Libraries

Dan Wu
School of Information Management
Wuhan University, Wuhan, Hubei
430072, China
woodan@whu.edu.cn

Nanhui Gu
School of Information Management
Wuhan University, Wuhan, Hubei
430072, China
gunanhui@qq.com

Daqing He
School of Information Sciences
University of Pittsburgh, Pittsburgh
PA 15260, USA
dah44@pitt.edu

ABSTRACT
Digital library, because of its resource demanding and other issues to solve, is an important application of multilingual information access (MLIA). However, the requirements of MLIA systems and applications are not typically addressed or assessed in our evaluations of digital libraries. This paper, therefore, aims to study the usages and expectations of MLIA in Chinese academic digital libraries. We conducted two surveys to study MLIA in current Chinese academic digital libraries and to get to know the users' real requirements for MLIA in Chinese academic digital libraries. The initial results offer thoughts on specific MLIA functions and insights on future digital library design and developments.

Categories and Subject Descriptors
H.3.7 [**Digital Libraries**]: User Issues

General Terms
Human Factors, Languages, Experimentation

Keywords
Multilingual Information Access, Digital Library, User Survey, Chinese Academic Libraries

1. INTRODUCTION
Multilingual information access (MLIA) studies the storage, access, retrieval and presentation of information in multiple languages. It is critical for the further integration of people and information at the global scale. One important application area of MLIA is digital library. Digital libraries (DL), which hold large scale digitalized resources, play important roles in media-rich life. Multi-media, multi-linguality and multi-culture are the three major characteristics of digital libraries [1]. As an integration of content and technology, digital libraries contain many MLIA related issues, which include multilingual resource management, multilingual portals, multilingual search, multilingual presentation of results, multilingual question and answering, multilingual text mining etc. Therefore, it is not surprise that increasing number of digital library services have realized the importance of MLIA. For example, the European Commission launched the i2010 Digital Libraries Initiative to enable multilingual access to the contents of Europe's national libraries [2].

However, MLIA systems have not been widely adopted except a few recent developments such as Google Translate. This is due, in part, to lack of demands in the marketplace, but also, in perhaps greater measure, to the special requirements that may be associated with MLIA applications – the requirements that are not typically addressed or assessed in our research evaluations. One such requirement is end-user support in MLIA systems as end users have greater needs for the translation or summarization of retrieved information [3].

Therefore, to offer insights to MLIA functions in future digital library applications, we conducted a series of studies on usages and expectations of MLIA in the current digital libraries, and our research focus has been on collecting the users' real requirements directly, and on Chinese academic digital libraries.

In the remainder of this paper, we will first review the related work on MLIA in digital library in section 2; then in section 3, we talk about our research questions and two sets of surveys in detail. Then, we will analyze the results of the two experiments and obtain answers to the research objectives in section 4. Finally, we will conclude with discussions and future works in section 5.

2. RELATED WORK
Research activities associated with MLIA in digital libraries can be divided into three aspects. The first one explores the framework of integrating MLIA with digital libraries. Oard [4] pointed out that users seeking information from a digital library could benefit from the ability to query large multilingual collections with a single language, and thesauri can help to address this challenge by facilitating controlled vocabulary search using terms from several languages. Maybury and Griffith [5] described an integrated environment for information analysts to examine very large multilingual collection. Chen [6] gave an overview of multilingual information access in digital library, in which National Palace Digital Museum was used as an example. Liu et al [7] created Arc, an OAI compliant federated digital library, and discussed how Arc can integrate an existing cross-language retrieval component. Pavani [8] studied Maxwell system Digital Library and identified basic functionalities and components for a multilingual digital library.

The second aspect is about research on multilingual information processing technology for constructing multilingual digital libraries. Bian and Chen [9] discussed cross-language information access to multilingual collections on the Internet. Wang et al [10] investigated the feasibility of exploiting the Web as the corpus source to translate unknown query terms for cross-language

information retrieval in digital libraries. Richardson and Fox [11] put forward a method using concept maps as a cross-language resource discovery tool for large documents in digital libraries.

The third one studies the usage of multilingual information resources in existing digital libraries or projects. A research team at the University of Maryland [12] designed the International Children's Digital Library which selects and processes books from different countries, and presents them in multiple languages simultaneously. Berkeley Public Digital Library [15] provides multilingual resources in eight languages, and has the multilingual catalog search and multilingual reference service. Comparing to the rest of the world, Europe pays more attention to the multilingual issues in digital libraries. Among 14 projects containing multilingual collections funded by the European commission under the Fifth Framework Programme [13], ETRDL project provided multilingual interface in six languages and multiple language text processing, and SCHOLNET was an extension of ETRDL with cross-language search functionality. ECHO was a project about film archives in four languages, and it had cross-language search via controlled vocabulary. MUCHMORE project was for CLIR in medical domain. MultiMatch project is a multilingual and multimedia search engine for cultural heritage [14], which has components for both document and query translation. The European digital library, museum and archive is a single access point to Europe's cultural heritage in multiple languages.

The reviewed literature shows many digital libraries are the products of collaboration from different countries which naturally produce bilingual or multilingual collections, and these digital libraries serve broader or global user communities with users speaking different languages, but many of them do not have multilingual search capabilities. More importantly, few studies have evaluated the multilingual collections of digital library from the user's perspective. The lack of user studies is surprising considering the increasing interest in digital library projects.

We identified similar situation inside China. The development of digital libraries in China has a shorter history, and few digital libraries have MLIA services. With China becoming more and more open to the world, it is necessary for Chinese researchers to access foreign language resources easily and to disseminate Chinese achievements to the rest of the world. Therefore, this paper aims to investigate the current usages of MLIA in Chinese academic digital libraries, and to elicit users' real requirements of MLIA in those digital libraries. Our research results will help Chinese digital libraries to develop and provide MLIA services that truly meet the users' needs.

3. RESEARCH DESIGN

3.1 Research Questions and Methodology

Our goal in this paper is to examine the usages and expectations of multilingual information access in Chinese academic digital libraries. We here have two research questions:

1) What are the current applications and usages of multilingual information access services in Chinese academic digital libraries?

2) What are the users' real requirements on the multilingual information access services in Chinese academic digital libraries?

We adopt survey as our data collection method to investigate the above research questions. We have conducted two surveys, and the first one sampled several representative Chinese digital libraries, and interviewed the managers of the libraries to obtain detailed information about the current usage of MLIA services. The other survey used a questionnaire to collect what users from different academic disciplines think of MLIA services in digital libraries.

3.2 Survey Design and Data Collection

3.2.1 Digital library survey

The goal of this survey was to investigate the usages of multilingual resources in Chinese digital libraries. We conducted a pilot study to examine the availability of multilingual resources in major public libraries and major academic libraries. We visited two libraries: the Shenzhen City Library which is a big public library, and the library of Wuhan Branch of Chinese Academy of Sciences which is an big academic library. Our investigation found that the amount of the multilingual resources is quite different between these two kinds of libraries (see Table 1).

Table 1. Results of the Pilot Multilingual Resources Survey

	Library of Wuhan Branch of Chinese Academy of Sciences	**Shenzhen Library**
Type	academic library	public library
% of total	80%	10%
Lang.	English (96%), Japanese (1%), French (1%), German (1%), Russian (1%)	English (98%), Japanese (0.5%), French (0.5%), German (0.1%), Korean (0.1%), Russian (0.1%), others (0.7%)

Table 2. The Six Selected Academic Digital Libraries

ID	**Digital Library (DL)**	**URL**
1	Wuhan University Library	http://lib.whu.edu.cn/
2	Huazhong University of Science and Technology Library	http://www.lib.hust.edu.cn/
3	Huazhong Normal University Library	http://lib.ccnu.edu.cn/
4	Huazhong Agricultural University Library	http://lib.hzau.edu.cn/
5	Zhongnan University of Economics and Law Library	http://lib.znufe.edu.cn/
6	Wuhan University of Technology Library	http://lib.whut.edu.cn/

Based on this pilot study, we decided to focus on the digital library services in Chinese academic libraries first. We selected the six academic digital libraries in Table 2 based on the following reasons: First, these are top research universities whose library systems would have multilingual information resources. Second, these universities cover different types of disciplines, such as science and technology, economics and law, agriculture, etc. Third, all these universities are located in Wuhan city, which

provide easy access from geographic point of view. We acknowledge that the digital libraries we studied in this paper are all from Wuhan city, however, based on the above reasons, they represent typical digital academic libraries in China.

Our survey to these Chinese academic digital libraries mainly consisted of highly-structured interviews with the managers of these libraries. The interviews were based on 11 questions that cover three major categories (See Table 3).

Table 3. Questions about the Usages of MLIA

Category	Question	Question Type
Basic Info.	Q1: What percentage of the digital resources is in foreign languages?	close
	Q2: What are the multiple languages of digital resources?	multiple choices
	Q3: What percentage of the budget is spent on foreign language resources?	close
	Q4: Who are the main users of the foreign language resources?	multiple choices
Usage of Multi-lingual Resources	Q5: How often are the multilingual digital resources visited?	Likert Scale
	Q6: Are you satisfied with the usages of the multilingual resources?	Likert Scale
	Q7: Do you think that the current multilingual digital resources can meet users' needs?	Likert Scale
Multi-lingual Services	Q8: Do you provide training for using the multilingual resources?	single choice
	Q9: Do you provide multilingual document delivery? How many papers per year?	single choice and close
	Q10: Do you provide multilingual search?	single choice
	Q11: What actions have you adopt to improve the usage of the multilingual resources?	multiple choices

3.2.2 Digital library user survey

To investigate the users' requirements to the multilingual services in digital libraries, we conducted the second survey by sending a questionnaire to potential digital library users in many academic disciplines. This questionnaire has 99 questions including 11 questions for user's demographic information, and the remaining 88 questions divided into six categories which include user behaviors, MLIA requirement motivations, multilingual information sources, multilingual services, multilingual searches, and multilingual interfaces (See Table 4). All these categories cover possible aspects where multilingual information resources could be used in digital libraries. Except the questions about users demographic information, all the remaining 88 questions use 5 level Likert scales (1 means "totally disagree", score 2 means "disagree", score 3 means "not sure", score 4 means "agree", and score 5 means "absolutely agree").

Table 4. Questionnaire about User Requirement of MLIA

Category	Survey Content	Question
0	User Demographic Info	11 questions
1	User Behaviors	Q1-Q15
2	MLIA Requirement Motivations	Q16-Q30
3	Multilingual Information Sources	Q31-Q52
4	Multilingual Services	Q53-Q69
5	Multilingual Searches	Q70-Q74
6	Multilingual Interfaces	Q75-Q88

The users we selected are mainly graduate students, teachers, or librarians from the selected universities in Table 2. We totally recruited 78 subjects to fill out the questionnaire. They represent different disciplines, which include information science, library science, computer science, telecommunication, electrical engineering, energy, biology, chemistry, environment, literature, foreign language, publishing, social science, art, etc. To encourage the users to carefully complete the questionnaire, we provide small gifts to them.

4. RESULT ANALYSIS

4.1 Current Usages of MLIA in Chinese Academic Digital Libraries

4.1.1 Basic information

The results for questions 1-3 in Table 3 are presented in Table 5. The identification number of the digital libraries presented in Table 5 is the same as that in Table 2. The results show that the percentage of the digital resources of foreign languages is pretty high in most of the DLs. Except the library with ID 5, all other DLs have more than 55% of foreign language resources. This means that multilingual digital resources are very important and abundant in those DLs. However, the coverage of languages is not wide. Only three DLs provide multilingual resources beyond English, the remaining three only have English collections. And to the percentages of the budget that the DLs spent on multilingual resources, only 3 DLs told us the amount. However, all these three DLs spend over 55% of their budget on foreign language resources. This means that if those multilingual resources are not effectively used, it is a waste of money.

Table 5. Result of Basic Information

DL ID as in Table 3	Q1	Q2	Q3
1	65%	7 languages	-
2	70%	English only	70%
3	60%	English only	-
4	70%	English only	80%
5	34%	5 languages	-
6	58.62%	5 languages	55.71%

The multiple choices of question 4 include foreign teacher, foreign student, Chinese teacher, Chinese student, and other staff. Our results show that both foreign and Chinese teachers and students are the main users of these multilingual resources, then are the staffs.

4.1.2 Usage of multilingual resources

For question 5-7 in Table 3, we used Likert Scale to test the usages of multilingual resources in those DLs. 1 to 5 scales represent from "totally not satisfied" to "very satisfied". As shown in Figure 1, only one DL (DL 4) has a very high visit count of its multilingual resources, all others have only average visit counts. The average satisfaction of the multilingual digital resources is only moderate throughout all DLs. And most managers think that their multilingual resources can generally meet the users' basic needs, but are not at very satisfied level. These results show that the usages of MLIA in these digital libraries are reasonable but not very satisfying.

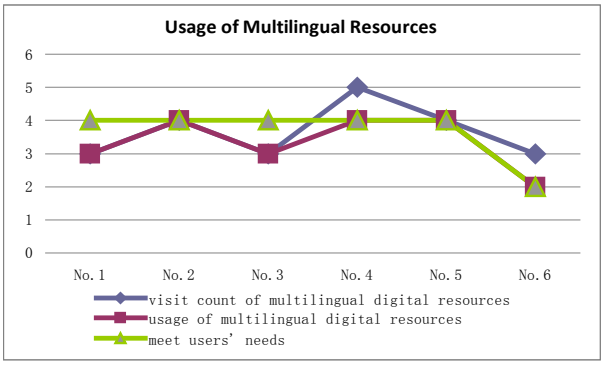


Figure 1. Usages of multilingual resources in the six DLs

4.1.3 Multilingual services

The remaining four questions in Table 3 investigate the multilingual services that those DLs provide. All six DLs have foreign language database training for their users. All of them offer multilingual document delivery services, and the average amount of paper delivered per year is about 4300, with the highest number as 12,000 per year at Wuhan University Library. This demonstrates that multilingual document is required by many users. As for multilingual search, only four of the six DLs have the function. The actions that the six DLs adopt to improve the usage of multilingual resources are as follows: (a) Resources Reorganization (3 DLs); (b) E-journals Navigation (4 DLs); (c) Information Retrieval Training (6 DLs); (d) Multiple Database Search (5 DLs); (e) Personalized Information Customization (4 DLs); (f) Online Translation Tool (None); (g) Cross-language Information Retrieval (1 DL in part of its database); (h) Others including promotion. This shows that the improvements are at general monolingual search and navigation, but the translation related multilingual search services like machine translation and cross-language information retrieval have not been widely adopted.

We can see from the above questions that most digital libraries can provide some basic multilingual services like multilingual portal, multilingual document delivery, multilingual resource training, etc. However, more advanced services, especially services like multilingual information access are missing. We think that this is the future emphasis that digital libraries should pay attention to.

4.2 User Expectations of MLIA in Chinese Academic Digital Libraries

To learn the users' requirements, we divided our questions into six categories as shown in Table 4. Here we will report the reliability, validity, average score and other results of each category. The statistic tool we used was SPSS 16.0. The measure we used to analyze the internal consistency reliability coefficient was Cronbach's Alpha, and the measure for validity was KMO (Kaiser-Meyer-Olkin) and Bartlett's Test. The values of the whole questionnaire and each category are shown in Table 6.

It can be seen that the alpha value of the whole questionnaire is 0.924, which means that the questionnaire is very reliable. As to individual category, only categories 1 and 5 have alpha values between 0.35 and 0.7, which means that the reliability is moderate. All other four categories have alpha values higher than 0.7. For validity, all the six categories have KMO values above 0.5, which demonstrate that the questionnaire is valid. Because of the space, we only present selected findings in the remaining of this section.

Table 6. Reliability and Validity of the Questionnaire

	Reliability	**Validity**			
	Cronbach's Alpha	KMO	Bartlett's Test of Sphericity		
			Approx. Chi-Square	df	Sig.
whole	0.924	-	-	-	-
category1	0.470	0.588	339.452	105	0.000
category2	0.771	0.674	411.538	105	0.000
category3	0.761	0.671	689.440	231	0.000
category4	0.915	0.837	803.737	136	0.000
category5	0.662	0.725	70.748	10	0.000
category6	0.902	0.841	602.553	91	0.000

4.2.1 User background

In this category, the character distributions of the 78 subjects are:

1) 56% of the subjects are male, 54% are female;
2) 2.5% are under 20 year old, 73.4% are between 21 to 30, 20.3% are between 31 to 40, 3.8% are over 40;
3) 36.7% received bachelor degree, 48.1% had master degree, 15.2% had doctor degree;
4) 26.6% are junior or senior professors, others are students;
5) 11.4% are from literature, art or history discipline; 32.9% are from social science; 17.7% are from science and technology; and 38% are from engineering;
6) Only 9% of them mastered two foreign languages, and others could only speak English as the second language;
7) Almost all of them had used online database, digital library, search engine, online public access catalogue (OPAC), and online translation tool before. The most frequently used tools were:
   online databases: CNKI and Wanfang Data;
   digital libraries or OPACs: Chaoxing DL, National DL, and university library catalogue;
   search engines: Baidu and Google

translation tools: PowerWord, Google Translate, and Lingoes

4.2.2 User behaviors

Questions Q1-Q15 investigate users' MLIA behaviors. The average score using 5 level Likert scales of this category is 3.13, and major interesting results are:

1) Users are not sure that they know DL very well.
2) Between multilingual books and journals in traditional paper based media and that in digital format, users are more willing to read multilingual digital resources.
3) Users are not sure that they have difficulty in searching or reading multilingual information.
4) Users are satisfied with their multilingual searches in the languages that they can understand.
5) Users have difficulty in searching for multilingual information in languages that they cannot understand, and they feel that they have the needs to access those information.
6) Users rely on translation tools to help them looking for information in the language that they cannot understand, but they are not satisfied with the translation quality.
7) When users received multilingual information in the languages that they cannot understand, they probably would give it up rather than asking help from friends or librarians.

These findings show that users are eager to and have needs to access to multilingual information, but they have problems in searching and receiving those information that they cannot understand. Translation tools are very helpful here and their quality is not satisfactory.

4.2.3 MLIA requirement motivations

Questions Q16-Q30 investigate users' motivations for accessing multilingual information. Among all reasons we listed, the average score is 2.92. Below are the reasons that received above 3.0 average scores. From highest to the lowest, they are:

1) I want to know the latest developments in other countries. (score: 3.74)
2) I need to write a literature review. (score: 3.64)
3) I need to finish an assignment. (score: 3.51)
4) I want to conduct a research that is novel. (score: 3.32)
5) I need to conduct my daily work. (score: 3.19)
6) I am asked by some friends for help. (score: 14)

The results show that the main reasons that motivate users to access multilingual information are related to research work rather than daily life. Therefore, we can conclude that MLIA in digital library is more necessary and important than in search engine.

4.2.4 Multilingual information sources

Questions Q31-Q52 examine how users use and evaluate multilingual information sources. The average score of this category is 3.61. The interesting finds are:

1) The sources from which users obtain academic information are search engines first, then digital libraries, then traditional libraries.
2) Users have used multilingual information when they search on search engines, digital libraries and library OPACs.
3) Users use Chinese digital libraries and search engines when they search for Chinese academic information, and they would use tools in other languages when they search for information in other languages.
4) Most users have tried Google Translate cross-language search engine before, and generally think that it is good. But few of them have tried Yahoo Babel Fish cross-language search engine.
5) Users are not satisfied with the multilingual information that current digital libraries, search engines, and traditional libraries provide.
6) Users are not sure whether they are satisfied with the ways that they access to multilingual information.
7) Users have a high expectation of a multilingual integrated digital library that has multilingual information access capabilities for languages that they are not familiar with.

These results show that when users need to find multilingual academic information, they mainly rely on search engines and digital libraries. But they are not satisfied with current multilingual tools, and know little about cross-language search engines. On the other hand, users all hope to use a multilingual integrated DL to access to information.

4.2.5 Multilingual services

Questions Q53-Q69 examine the multilingual services that digital libraries should offer. The average score of this category is 3.98. We identified a set of services in the questionnaire and below are the top services that have the average scores above 4.0:

1) provide clustered multilingual information based on subject or discipline (score: 4.29)
2) provide professional term translation (score: 4.27)
3) provide translation assistance for any language (score: 4.18)
4) provide abstract translation for any language (score: 4.18)
5) provide multilingual expert clustering (score: 4.17)
6) allow users to add term translations (score: 4.12)
7) allow users to customize multilingual RSS feeds (score: 4.10)
8) allow users to correct the wrong translation (score: 4.09)
9) allow users to add tag to multilingual resources (score: 4.06)
10) provide language statistics (score: 4.05)
11) provide multilingual search (score: 4.01)

We can see that users really want some specific MLIA services in their DLs. What they want the most are term translation and abstract translation for any language rather than full-text translation. And they also need clustering functions to organize the multilingual resources as well as some interactive services.

4.2.6 Multilingual search

Questions Q70-Q74 reveal the multilingual search functions that DL should offer. The average score of this category is 3.90. Below are the functions whose scores are higher than 4.0:

1) allow users to search in their native language on multiple language documents (score: 4.29)
2) provide translation assistance for users to choose correct translation of the terms (score: 4.22)
3) translate the abstract of the retrieved documents to the users' native language (score: 4.05)

Therefore, cross-language information retrieval is necessary for the users. The users also want to translate the retrieved results back to their native language. They need assistance in term translation for translation disambiguation.

4.2.7 Multilingual interface

Questions Q75-Q88 ask for the ideal multilingual interface for DLs. The average score of this category is 3.89. Here are the

features received the average score above 4.0. From the highest to the lowest, they are:

1) DL should classify the retrieved multilingual results based on language. (score: 4.12)
2) For those non-popular language information, DL could use English to describe. (score: 4.08)
3) For term or sentence translation, DL should provide concept explanations for helping to select correct translation. (score: 4.05)
4) For term or sentence translation, DL should provide translation probability for helping to select correct translation. (score: 4.04)
5) DL should offer multilingual OPAC. (score: 4.01)

These findings show that users prefer the DL interface that can provide more assistance for translation.

5. CONCLUSION

In this paper, we conducted two surveys to investigate what are the usages and expectations of multilingual information access in digital library. The results provide answers to the two research questions we proposed:

1) Multilingual resources which cost lots of money to construct are very important and abundant in the digital libraries that we studied, and most of these digital libraries can provide some basic multilingual services. However, the usage of MLIA in these digital libraries has not been very satisfactory, and more advanced services such as translation and multilingual information access are critically needed.

2) The DL users are eager to access to multilingual information, but they have problem in finding the information they cannot understand. Therefore, they need helpful translation tools and multilingual DLs which integrate tools that users frequently use to access to multilingual information. Furthermore, users prefer to use interactive DL interfaces that can provide more MLIA interactions, such as term and abstract translation functions for any language, clustering functions, translation assistance, and translating the retrieved results back to their native language.

Our future work includes: 1) We will confirm our findings through studying more Chinese academic digital libraries. 2) We will study further the difference of the usages and expectations of MLIA in Chinese digital libraries and that of digital libraries in other countries. 3) We will propose methods for better design of Chinese digital libraries, especially at the MLIA services. 4) We will compare the two survey results and articulate how the findings of our study can help the library managers to improve their multilingual services.

6. ACKNOWLEDGEMENT

This work is partially supported by National Social Science Foundation of China under the agreement 09CTQ026, Social Science Foundation of Chinese Education Ministry under the agreement 09YJC870022, and Youth Social Science Foundation of Wuhan University under the agreement 08QNXM39.

7. REFERENCES

[1] Borgman, C.L. Multi-Media, Multi-Cultural, And Multi-Lingual Digital Libraries, Or How Do We Exchange Data in 400 Languages? D-Lib Magazine, June 1997, vol. 3, no. 6. http://www.dlib.org/dlib/june97/06borgman.html [01-11-2009].

[2] Gey, F.C., Kando, N., Lin, C., Peters, C. New Directions in Multilingual Information Access: Introduction to the Workshop at SIGIR 2006. The 5th Workshop on Important Unresolved Matters, a Workshop of SIGIR 2006. August 2006, pp. 1-2.

[3] Evans, D.A. From R&D to Practice – Challenges to Multilingual Information Access in the Real World. The 5th Workshop on Important Unresolved Matters, a Workshop of SIGIR 2006. August 2006, pp. 3.

[4] Oard, D.W. Serving Users in Many Languages. D-Lib Magazine. December 1997, vol. 3, no. 12. http://www.dlib.org/dlib/december97/oard/12oard.html [01-11-2009].

[5] Maybury, M., Griffith, J., Holland, R., Damianos, L., Hu, Q. and Fish, R., Virtually Integrated Visionary Intelligence Demonstration, MITRE technical papers, http://www.mitre.org/work/tech_papers/tech_papers_05/05_0140/05_0140.pdf [01-11-2009].

[6] Chen, H.H. Multilingual Information Access in Digital Library. International Conference on e-Education 2004: Review and New Perspectives, June 24-25, 2004, Macao.

[7] Liu, X., Maly, K., Zubair, M., Hong, Q., Xu, C. An OAI Compliant Federated Digital Library. Scientific Commons. 2008. http://en.scientificcommons.org/42412620 [01-11-2009].

[8] Pavani, A.M.B. A Model of Multilingual Digital Library, 2001. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000300010 [01-11-2009].

[9] Bian, G.W., Chen, H.H. Cross-Language Information Access to Multilingual Collections on the Internet. Journal of the American Society for Information Science. 51 (3): 281-296, 2000.

[10] Wang, J.H., Teng, J.W., Cheng, P.J., Lu, W.H., Chien, L.F. Translating Unknown Cross-Lingual Queries in Digital Libraries Using a Web-based Approach. Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries (JCDL'04), 2004: 108-116.

[11] Richardson, R., Fox, E.A. Using Concept Maps as a Cross-Language Resource Discovery Tool for Large Documents in Digital Libraries. Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries (JCDL'05). 2005: 415.

[12] Hutchinson, H.B., Rose, A., Bederson, B.B. Weeks, A.C., Druin, A. The International Children's Digital Library: A Case Study in Designing for a Multi-Lingual, Multi-Cultural, Multi-Generational Audience. Information Technology and Libraries. 2005, vol. 24, no. 1, pp. 4-12.

[13] http://ec.europa.eu/research/fp5.html [01-11-2009].

[14] http://ec.europa.eu/research/fp6/index_en.cfm [01-11-2009].

[15] http://www.berkeleypubliclibrary.org/services_and_resources/multilingual_resources.php [17-11-2009]

Exploring the Further Integration of Machine Translation in Multilingual Information Access

Daqing He
School of Information Sciences
University of Pittsburgh, Pittsburgh PA 15260, USA
dah44@pitt.edu

Dan Wu
School of Information Management
Wuhan University, Wuhan 430072, China
woodan@whu.edu.cn

ABSTRACT

Machine Translation (MT) has been identified as a very important related technology for Multilingual Information Access (MLIA). Over the past decade, the usages of MT in MLIA are still largely concentrated on its capabilities for document translation, selection and examination. In this paper, by using a common evaluation framework, we explored the applications of MT in several unexplored or underexplored areas, which include query translation, relevance feedback, and out of vocabulary term resolution. Our experimental results demonstrate the unique contributions that MT can provide in those areas, and at the same time raise more interesting questions about how MT can be optimally integrated with MLIA.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – Search Process

General Terms

Design, Experimentation, Languages, Performance

Keywords

Multilingual Information Access, Machine Translation, Query Translation, Relevance Feedback, Out-Of-Vocabulary Term.

1. MACHINE TRANSLATION IN MLIA

With vast amount of multilingual information on the Web and the ability to read the results, it is natural for users to issue queries in one language, and access documents in other language(s). This so called Multilingual Information Access (MLIA) has been an active research area for more than a decade. Translation has played a very important role in MLIA. Although there are techniques for achieving MLIA without actually involving translation [1], most MLIA techniques often rely on translation methods to cross the language barriers between a query and the documents. Depends on whether it is the query, the documents, or both that are translated, we have document translation based MLIA (DT-MLIA), query translation based MLIA (QT-MLIA) and interlingual MLIA (IL-MLIA). Many resources have been exploited for the translation task, among which the most commonly used are machine-readable dictionary (MRD), parallel or comparable corpora, and machine translation (MT). MRD is probably the most commonly used in experiment setting, especially for translating queries [2]. However, considering that several companies such as Google and Yahoo are actively prompting their multilingual MT services on the Web, MT is probably the most easily accessible translation resource among the above three on the Web between commonly used language pairs. Therefore, it is important to examine the usages of MT in MLIA.

We believe that from the view point of MLIA, MT can be viewed as either a component of MLIA or as one of the major translation resources for MLIA. This motivated us to look at the usages of MT in MLIA in the steps of QT-MLIA such as query translation, relevance feedback, interactive MLIA, and out-of-vocabulary (OOV) term translation. The reason for concentrating on QT-MLIA is because it is query translation that makes MLIA different to monolingual information access. QT-MLIA reveals its translation process to the users so that the users can feel and build up more control of the search process. To the wide range of MLIA users, it is probably QT-MLIA that they will most interact with if they want to perform MLIA. Our goal is to obtain more insights about the wide range usages of MT in MLIA, and to help us and the community to identify promising future directions for both MT and MLIA.

The remainder of this paper is organized as follows. We discuss in detail our research topics of MT in MLIA and the corresponding experiment settings in Section 2. And in Section 3, we will present our insights to the usages of MT in MLIA, and conclude with brief highlights of our future work.

2. RESEARCH TOPICS AND EXPERIMENTS

2.1 General Experiment Settings

There are two research angles in our experiments and the discussions here. The first one examines the effectiveness of different query lengths on the studied techniques. The motivation is that Web search queries are often short, but queries in TREC like evaluation frameworks are often much longer. Previous studies show techniques often are more suitable for certain types of query lengths. Our experiments conducted under different query lengths, therefore, will provide more insights about the applicability of the techniques. The second angle is about technique integration. Over the years of active research, there have been many techniques and methods developed for MLIA. One insight obtained in the literature is that different techniques and methods often can be combined to obtain further

improvement. Therefore, when possible, we will examine the integration of complementary techniques in our studies.

All studies reported in this paper were performed on the same experiment environment. This not only helped to simplify the experiment design, but also made it possible to compare results across several studies.

Our experiments were performed between English queries and Chinese documents. The test collection contains documents from TDT4 and TDT5 Multilingual corpora. All the documents in the collection are news articles in the time period of 2000 to 2003 from several news agencies including Xinhua News Agency, Zaobao News Agency, China Broadcasting System, etc. The two TDT collections contain 83,627 Chinese documents and corresponding number of machine translation documents generated by ISI MT system. The collections also contain 306,498 English documents from several English news agencies at the same time period as Chinese documents.

We selected 44 TDT English topics and manually translated them into Chinese for monolingual Chinese search. These topics were also converted into TREC topic style with title, description and narrative fields for our study of the effect of different query lengths. Queries were automatically extracted from the topics with short queries containing titles only (T query), medium queries with title and description fields (TD query), and long queries with all the three fields (TDN query). The average length of the queries were: T query (4 terms), TD query (27 terms), and TDN query (127 terms).

The bilingual MRD used for our dictionary-based MLIA was an English-Chinese lexicon generated from a parallel bilingual corpus automatically [3]. The dictionary contains 126,320 English entries with translation probabilities for each Chinese translation alternative. During the translation of the queries with the MRD, to remove low probability translations which often are noises, a fixed threshold called Cumulative Probability Threshold (CPT) was selected. A threshold of 0 corresponds to the using the single most probable translation (a well-studied baseline), and a threshold of 1 corresponds to the use of all translation alternatives in the dictionary. In order to improve the coverage of the dictionary as much as possible, we adopted the back-off translation strategy [4] during the translation of the query terms.

In MRD based MLIA, we adopted a named entity translation component based on information extraction (IE) techniques. The NE component is designed to provide two functions in the MRD based query translation. The first one is to identify NEs in a given text, which could be queries, documents, or any parts of queries and documents. The function was provided by the NYU English and Chinese HMM-based name taggers trained on several years of ACE (Automatic Content Extraction) corpora. Both name taggers can identify names such as Person, Geo-Political Entity (GPE), Location, Organization, Facility, Weapon and Vehicle, and achieve about 87%-90% F-measure on newswire [5].

If translation enhancement (will be presented in section 2.3) was involved in the experiment, we selected TE-TWA method. The CPT threshold was 0.5, and λ was 0.5. Both of the two values were obtained via training [6]. If query expansion (QE) was involved in the experiments, we used the Indri's build-in PRF module which is based on Lavrenko's relevance model [7] . Depends on whether the QE is performed before and/or after query translation, we have pre-translation, post-translation and combined QE. The parameters in QE were set as top 20 terms from top 20 returned documents. The weight between original query and expanded terms is 0.5. This was based on our previous exploration of the parameters in Indri.

Unless mentioned specifically, the measure used was Mean Average Precision (MAP) over a ranked list. This measure is a commonly used evaluation measure in IR field. Statistical significance tests used in all our experiments were two tailed paired samples t-test, and we used p-value < 0.05 as the threshold for the statistically significance.

2.2 Topic 1: MT for Query Translation

The core step in QT-MLIA is the translation of queries, and MT can be integrated for translating queries. However, the effectiveness of using MT for translating queries comparing to MRD based methods is uncertain in previous studies [8]. Recently, both MT and MLIA have experienced rapid integration of statistical based language models and resources into their handling of translations. Statistical MT has become the state of the art for MT, and even some commercial MT systems such as Google Translate are statistical MT systems. Translation probabilities are widely used in MLIA for handling translation ambiguities or are even built as a part of the statistical language modeling for MLIA [2, 9, 10]. One important insight gained in MLIA from the usage of translation probabilities is that choosing multiple translations with their probabilities is a superior method than choosing only the top best translation. This insight actually to some degree argues against the current usage of MT output, which contains only one best translation for query terms or documents.

Therefore, the objective of this research topic is to examine again the effect of MT in query translation. We concentrate on using an out-of-box commercial MT system – Google Translate -- for the task. Our motivation is that if commercial MT systems have demonstrated their capabilities in translating queries, maybe effective MLIA capabilities can be easily constructed even by layman users. The users do not have to go through the steps of obtaining MRDs with high quality translation probabilities in order to perform MLIA. What they need is an online MT system.

Table 1: The MAP results of MT-based and MRD-based runs (* indicates that the improvement is statistically significant between MT Plain and MRD Base)

Run ID	T	TD	TDN
Mono Base	0.4739	0.5817	0.6215
MT Plain	0.4446*	0.5536*	0.6170*
MT QE-PreTrans	0.4922	0.5443	0.5580
MT QE-PostTrans	0.5284	0.6031	0.6292
MT QE-Combine	0.5604	0.5833	0.6001
MRD Base	0.3336	0.4251	0.4701
MRD QE-PreTrans	0.3714	0.4377	0.4477
MRD QE-PostTrans	0.4118	0.5080	0.5182
MRD QE-Combine	0.4415	0.5007	0.5170

The research questions associated with this study are: 1) when no other technique is integrated, can MT based query translation

perform comparably to monolingual search or to MRD based query translation; 2) if performance enhancement technique such as query expansion (QE) based on pseudo relevance feedback (PRF) is used, would MT based query translation still work; and 3) is query length a factor that affects MT based query translation? When using the MT system for query translation in the experiments, we entered the whole query into the MT system at once.

The experiment revealed some interesting results. As shown in Table 1, when there is no RF performed, MT based run "MT-Plain" performed really well. Its performance values under three different query lengths were between 94% to 99% of monolingual run "Mono Base". This is comparable to the state of the art MRD based MLIA performance, which integrated many performance enhancement techniques. In addition, the MT based runs obtained significantly improvement over the plain MRD baseline "MRD Base" under all three different lengths of queries.

Query length affects MT based query translation methods. MT based query translation works the best with long queries. The performance of "MT Plain" can reach to 99% of that of "Mono Base". However, it is interesting to see that the superiority of MT method is shrinking along with the increasing of query length comparing to "MRD Base". Maybe this does not imply that MT method is not good for long queries, it probably just means that MRD method performs better with long queries too. The longer the queries are, the more information can be used in the MRD method to compensate the impact of translation ambiguities. Because the MT method worked well over short queries too, this helps to remove the worry that MT does not work when not much context is available for translation.

We know from previous studies in the literature, QE in general helps MRD-based methods. Our results in Table 2 show that QE was helpful to MT based method too. With QE, MT-based method was at least 90% of monolingual performance. Some QE methods even helped MT method to achieve over 100% of monolingual performance ("MT QE-Combine" at T queries achieved 120%). Therefore, with the help of a simple QE technique, most MT-based MLIA runs actually outperformed the corresponding monolingual runs. All these results confirm again that MT is useful for translating queries in MLIA.

Table 2: Comparison of QE methods for MT-based MLIA (* indicates that the improvement is statistically significant)

Run ID		Perc. Of Mono Base	Impr. over MT Plain
T	MT QE-PreTrans	103.86%	10.71%*
	MT QE-PostTrans	111.50%*	18.85%*
	MT QE-Combine	118.25%*	26.05%*
TD	MT QE-PreTrans	93.57%	-1.68%
	MT QE-PostTrans	103.68%	8.94%*
	MT QE-Combine	100.28%	5.36%
TDN	MT QE-PreTrans	89.78%*	-9.56%*
	MT QE-PostTrans	101.24%	1.98%
	MT QE-Combine	96.56%	-2.74%

Again, query length affects the performance of QE in MT based query translation. As shown in Table 2, similar to that in MRD-based MLIA, it is often the post-translation QE and/or the combined QE that are the best methods among the three MLIA QE methods. In fact, post-translation QE ("MT QE-PostTrans") seems to be helpful whatever the query length is, whereas pre-translation QE ("MT QE-PreTrans") only works for short queries and "MT QE-Combine" does not work for long queries. The reason why the longer queries don't work is that the translation resource is very good (the MT Plain condition is already with very high performance), thus there is less need for QE techniques to combat the translation failures.

Overall, not only MT is a reasonable query translation method in MLIA, but also it is a better tool for query translation than MRD, no matter what the query length is and whether QE is used.

2.3 Topic 2: MT for Relevance Feedback in MLIA

Revealed in the interactive track of Cross-Language Evaluation Forum experiments (such as in [11]), and demonstrated by Google's cross-language search engine inside Google Translate, a user who performs QT-MLIA tasks would not only need queries to be translated so that the retrieval can be performed, but also need the returned documents to be translated back so that the user can perform relevance judgment or document examination. This MLIA process with two translation steps imposes some interesting problems and opportunities about MT in MLIA, especially the MLIA relevance feedback (RF) techniques based on MT.

Translation Enhancement (TE) is one such relevance feedback technique in MLIA [6]. TE utilizes the MT outputs of relevant documents for improving query translation. It views that users' relevance judgments are performed on these MT outputs rather than on the original returned documents in another language. This understanding helps to treat the user confirmed relevant documents and their MT outputs as a small parallel corpus. By using a word alignment tool GIZA++ [12], individual words inside the relevant documents and their MT translations can be connected. Instances of translations of query terms and their probabilities can be extracted and integrated with the corresponding translation information in the original dictionary. This TE method is called Translation Extraction with Word Alignment (TE-TWA).

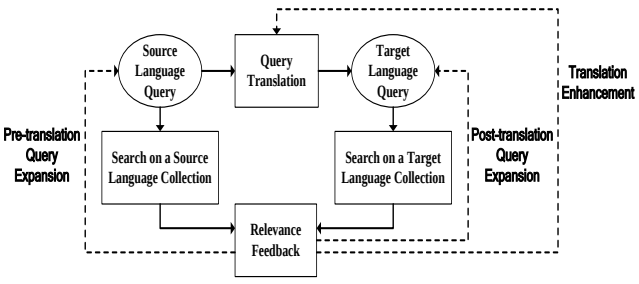


Figure 1: Translation Enhancement and Query Expansion (pre and post-translation) in MLIA

As shown in Figure 1, TE and various QE methods are applicable at different parts of the MLIA process, and they use different aspects of the relevant documents. The applications of their

relevance feedback results are different too: QE results are integrated into the query, and TE results modify the query translation, respectively. Therefore, two important research questions to be answered are: 1) what is the effect of TE-TWA method; and 2) what is the effect of combining TE and QE in MLIA?

As shown in Table 3, our experiment results show that TE approach "TE-TWA" performed better than the plain MLIA baseline run "MRD Base". The differences were statistically significant with all three types of queries. This demonstrates that TE is a valid and effective RF technique for improving MLIA performance. Comparing to the higher MLIA baseline "MRD-QE"1, the TE run "TE-TWA" outperformed "MRD-QE" when the queries were TD and TDN. However, only the improvement obtained with TDN queries was statistically significant. With TE alone, the MLIA performance cannot outperform the corresponding monolingual baselines "Mono-Base". However, "TE-TWA" matches 93.61% of the effectiveness of "Mono-Base" under the TDN queries, which is close to the state of the art MLIA performance.

Table 3: Comparison of TE-TWA with several baselines (* indicates that the improvement is statistically significant)

Run ID		MAP (Perc. of Mono-Base)	Impr. over MRD-Base
T	Mono Base	0.4739(100%)	+42.06%*
	MRD Base	0.3336(70.39%)	-
	MRD QE-Combine	0.4415(93.16%)	+32.34%*
	TE-TWA	0.3992(84.24%)	+19.66%*
TD	Mono Base	0.5817(100%)	+36.84%*
	MRD Base	0.4251(73.08%)	-
	MRD QE-PostTrans	0.5080(87.33%)	+19.50%*
	TE-TWA	0.5340(91.80%)	+25.62%*
TDN	Mono Base	0.6215(100%)	+32.21%*
	MRD Base	0.4701(75.64%)	-
	MRD QE-PostTrans	0.5182(83.38%)	+10.23%*
	TE-TWA	0.5818(93.61%)	+23.76%*

Query length affects the performance of RF techniques. As shown in the last column of Table 3, QE obtained the highest MAP improvement over the corresponding baselines "MRD-Base" with short T queries, and the improvement decreased with longer TD and TDN queries. However, the effect is different for TE runs. "TE-TWA" performed better with longer TD and TDN queries, but less so to that of the short T queries. Therefore, it seems that query length has different effects on TE and QE. This is another motivation to combine these two RF methods.

The combination of TE and QE ("TEQE") achieved comparable results to the monolingual baseline "Mono-Base" for all three types of queries (see Table 4). In the case of T and TD queries, "TEQE" even exceeded "Mono-Base". Of course, these runs still cannot outperform the higher monolingual baseline "Mono-QE". "TEQE" run also significantly outperformed TE only run under T and TD queries, and it significantly outperformed QE only run under TD and TDN queries.

Table 4: Comparison of combining TE and QE (TEQE) to several baselines and to TE or QE alone (* indicates that the improvement is statistically significant)

TEQE	MAP			
	MAP	Perc. of Mono-Base	Impr. over TE-TWA	Impr. over MRD-QE
T	0.4748	100.19%	+18.94%*	+7.54%
TD	0.5905	101.51%	+10.58%*	+16.24%*
TDN	0.5972	96.09%	+2.65%	+15.25%*

Another interesting point is that "TEQE" showed stable performance with all three types of queries. Different to TE that works better with long queries, and QE that works well with short queries but is losing performance with long queries, the combined run performed consistently comparable to the monolingual run with all three types of queries. It seems that the combination helped to use one's advantages to overcome the limitations of the other. Therefore, we can conclude that it is beneficial and effective to combine TE with QE.

2.4 Topic 3: MT for OOV Terms

Both MT and MLIA have to face out-of-vocabulary (OOV) terms. OOV terms refer to the words whose translations are not available in the translation resources such as MRDs [13]. In MRD based MLIA, it is beneficial to use dedicated data mining and information extraction methods for obtaining high quality translations for named entities [14], which are the most important and most common type of OOV [15]. We believe that a well designed MT system would be helpful in resolving OOV terms. This is not only true in query translation, but also true in TE method in relevance feedback for MLIA.

The research questions therefore are: 1) in MT for query translation, would MT which has its own handling of OOV terms generate comparable result to an MRD based query translation method that has a dedicated OOV module; and 2) can MT supported TE method helps in resolving OOV terms?

Table 5: The comparison between MT method and MRD method with dedicated OOV module (* indicates that the improvement is statistically significant over "MRD Base", † indicates that the improvement is statistically significant over "MRD NE Enhance")

Run ID	T	TD	TDN
Mono Base	0.4739	0.5817	0.6215
MT Plain	0.4446†	0.5536	0.6170†
MRD Base	0.3336	0.4251	0.4701
MRD NE Enhance	0.3934*	0.5034*	0.5563*

In the studying of MT for query translation, we observed from Table 5 that a dedicated NE module for MRD based query

[1] "MRD-QE" refers to corresponding "MRD QE-Combine" or "MRD QE-PostTrans" based on query length.

translation indeed helped the system performance. The run "MRD NE Enhance", which has a dedicated NE translation module, significantly outperformed "MRD Base" in all three types of queries. However, its performance was still inferior to MT based method "MT Plain", and in the case of T and TDN queries, the differences were statistically significant. This indirectly indicates that MT has its capability of handling OOV terms.

When examining individual topics, we noticed that there were many named entities in the topic statements, and Google Translate system handled them well. There were still a few NEs that Google Translate system cannot handle. But they were also OOV terms for the MRD even with the NE module. So they did not make obvious difference between the MT methods and the MRD methods.

Table 6: OOV terms and their translations found by TWA (# after the No indicates that the translation is wrong)

No.	Topic ID	OOV Term	Translations found by TE-TWA
1	55087	Bingol	宾格尔省
2	55087	diyarbakir	迪亚巴克尔
3#	40007	Garner	还
4	55087	kandilli	坎迪利
5#	55029	karolinska	推动/科技
6	55179/55127	Kumba	昆巴
7	41025	montesinos	蒙特西诺斯
8	40037	morariu	莫拉留
9	41012	ouattara	瓦塔拉
10	55181	Qurei	库赖
11	41025	vladimiro	弗拉迪米罗

In the experiments of TE, we found that during the process of extracting translation information from the small parallel corpus built from relevant documents and their MT translations, the TE method "TE-TWA" can identify the translations for some OOV terms with the help of word alignment information. Table 6 shows the 11 OOV terms and their translations found by "TE-TWA" method. Almost all these terms are NEs of people names, locations, etc. This is consistent with the finding by [13]. Therefore, it becomes an important advantage for "TE-TWA" (thus for MT) to be able to identify translations for some OOV NEs. Of course, as shown in In the studying of MT for query translation, we observed from Table 5 that a dedicated NE module for MRD based query translation indeed helped the system performance. The run "MRD NE Enhance", which has a dedicated NE translation module, significantly outperformed "MRD Base" in all three types of queries. However, its performance was still inferior to MT based method "MT Plain", and in the case of T and TDN queries, the differences were statistically significant. This indirectly indicates that MT has its capability of handling OOV terms.

When examining individual topics, we noticed that there were many named entities in the topic statements, and Google Translate system handled them well. There were still a few NEs that Google Translate system cannot handle. But they were also OOV terms for the MRD even with the NE module. So they did not make obvious difference between the MT methods and the MRD methods.

Table 6, some of the found translations are wrong (such as the translations for Nos. 3 and 5), which were the results of word alignment errors. However, the fact that majority found translations are correct indicates that it is reasonable reliable to use MT based TE method for resolving OOV terms.

In summary, based on the results obtained from our studies on MT for query translation, we can see that because MT systems often has its way of handling OOV terms, it often can overcome most OOV problems faced in MLIA process. Unlike a dedicated OOV term resolution module is needed in MRD based method, MT for query translation does not critically need an OOV module. Still its performance is either comparable or significantly better than MRD enhanced by an OOV module.

Based on the results from our studies on MT for RF in MLIA, it also makes sense to resolve OOV terms by performing translation enhancement. The extracted translation information contains possible solutions to some OOV terms. Because of the high quality of MT outputs, the translations of OOV terms obtained this way are in high quality too.

3. DISCUSSIONS

Overall, if the MT outputs of the whole collection are available, DT-MLIA is probably the simplest and cost-effective method. Literature has shown that the performance of DT-MLIA is among the best of the various MT usages in MLIA. Of course, relevance feedback techniques can still be applied to further enhance the results. However, QT-MLIA is better than DT-MLIA to give more transparency to MLIA searches. In this case, MT is a very effective method for translating queries. Considering that commercial MT systems are easily accessible online for many major language pairs, this is a very simple way of building MLIA capabilities. Query translation based on MT can by itself achieve comparable results to monolingual baseline, which is at the state of the art MLIA performance. If relevance feedback techniques like QE are applied, the final results would be close to performing QE on monolingual searches.

In addition, with MT's relatively full coverage of terms in general domain, QT-MLIA using MT for query translation does not have to have a dedicated NE module to handling OOV terms. This further simplified the design of MLIA system without sacrificing the retrieval performance. If MT output is readily obtainable, it is useful for resolving OOV terms too, and TE can achieve that.

In MLIA, besides QE as a basic relevance feedback method, we also demonstrate that TE, which is performed on the identified relevant documents and their MT outputs, is certainly a valid relevance feedback method too. Both QE and TE can achieve results close to but not exceed to the monolingual plain search. However, the most interesting property of them is that the combination of them can generate a much more robust MLIA relevance feedback mechanism that is capable of handling queries with different lengths.

It is better that MT should be applied before relevance feedback. The different effects obtained from different QE methods in MT based query translation raise an interesting insight about combining MT based query translation and QE. Because pre-

translation QE adds many expansion terms before MT can translate the queries, adding too many words that are not part of a sentence could hurt the quality of MT even though some of the words are truly relevant. This is true especially when the original queries already have many words to work on (such as in TD and TDN). So if QE is integrated with MT in query translation, it is better that QE is performed after MT.

At the same time, our studies also demonstrate that further integration between MT and MLIA is needed. For example, our current TE study still needs to apply word alignment between MT outputs and their original documents. This is expensive and should not be necessary. Statistical MT systems should be able to provide information similar to word alignment, or even better phrase level alignment, as part of its translation output. Mining from there is a much better approach for TE. Another example is that MT for query translation still gives out only one best translation. When the translations of the queries are simple and straightforward, this is not problematic. However, when the hypotheses of the translations of the query terms are at low confidence or quality, it is actually better for MT system to give out n-best translations. But current MT systems do not provide such output even though they are capable of doing that.

Finally, we acknowledge the limitations of our studies. Only one language pair and translation direction was used. All our experiments were performed on TDT collections, which are just one of the several available MLIA test collections. Many findings and insights could need further testing in other collections, and the document collections were quite comparable, which makes it more likely that pre-translation QE will be effective.

4. CONCLUSION

Machine Translation (MT) has been identified as a very important related technology for Multilingual Information Access (MLIA). In this paper, the primary goal is to test how MT can be used in MLIA process and what effects MT will bring to certain aspects of MLIA process. From multiple aspects of MLIA process, our studies demonstrate that MT can be applied at the many places in the whole MLIA process. Although MT in DT-MLIA certainly is a simple and cost-effective way of integrating MT into MLIA, many other important aspects of MLIA can certainly benefit from MT too.

Our future work on integrating MT in MLIA is in the following areas. First, applying word alignment on the returned documents and their machine translations is actually not the optimal approach. Current statistical MT systems can generate outputs with word level alignment information. We will explore the usage of such information in TE experiments. Second, improving translation relationship at word level certain helps, but modern phrase MT systems can give us many phrase level translation information. We can integrate the phrase translation generated by MT system into MLIA.

ACKNOWLEDGEMENT

This work is partially supported by National Science Foundation of USA under the agreement NSF/IIS 0704628, National Post-doctor Foundation of China under the agreement 20090451078, and Social Science Foundation of Wuhan University under the agreement 09ZZKY097.

REFERENCES

[1] Oard, D.W. and A.R. Diekema, *Cross-Language Information Retrieval*, in *Annual Review of Information Science and Technology*, B. Cronin, Editor. 1998, American Society for Information Science. p. 223-256.

[2] Darwish, K. and D.W. Oard. *Probabilistic Structured Query Methods*. in *Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2003: ACM Press.

[3] Wang, J. and D.W. Oard. *Combining Bidirectional Translation and Synonymy for Cross-language Information Retrieval*. in *Proceedings of the ACM SIGIR 2006*. 2006.

[4] Resnik, P., D. Oard, and G. Levow. *Improved Cross-Language Retrieval using Backoff Translation*. in *First International Conference on Human Language Technologies*. 2001.

[5] Grishman, R., D. Westbrook, and A. Meyers. *NYU's English ACE 2005 System Description*. in *ACE 2005 Evaluation Workshop*. 2005.

[6] He, D. and D. Wu. *Translation enhancement: a new relevance feedback method for cross-language information retrieval*. in *Proceeding of the 17th ACM conference on Information and knowledge management*. 2008. Napa Valley, California, USA: ACM.

[7] Lavrenko, V. and W.B. Croft. *Relevance-based Language Models*. in *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2001: ACM Press.

[8] Kraaij, W., *TNO at CLEF-2001: Comparing Translation Resources*, in *Evaluation of Cross-Language Information Retrieval Systems: Second Workshop of the Cross-Language Evaluation Forum, CLEF 2001 Darmstadt, Germany, September 3–4, 2001 Revised Papers*, C. Peters, et al., Editors. 2001, Springer Berlin: Heidelberg. p. 141-162.

[9] Lavrenko, V., M. Choquette, and W.B. Croft, *Corss-Lingual Relevance Models.* Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2002: p. 175-182.

[10] Xu, J. and R. Weischedel, *A Probabilistic Approach to Term Translation for Cross-Lingual Retrieval;*, in *Language Modeling for Information Retrieval*, W.B. Croft and J. Lafferty, Editors. 2003.

[11] Oard, D.W. and J. Gonzalo. *The CLEF2001 Interactive Track*. in *the Cross-Language Evaluation Forum (CLEF) 2001* 2001.

[12] Och, F.J. and H. Ney, *A Systematic Comparison of Various Statistical Alignment Models.* Computational Linguistics, 2003. **29**(1): p. 19-51.

[13] Demner-Fushman, D. and D.W. Oard. *The Effect of Bilingual Term List Size on Dictionary-Based Cross-Language Information Retrieval*. in *36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track 4*. 2003. Hawaii.

[14] Ji, H., et al. *NYU-Fair Isaac-RWTH Chinese to English Entity Translation 07 System*. in *NIST ET 2007 PI/Evaluation Workshop*. 2007.

[15] Thomas Mandl and C. Womser-Hacker. *The Effect of Named Entities on Effectiveness in Cross-Language Information Retrieval Evaluation.* in *ACM SAC'05*. 2005. Santa Fe, NM.

To Keep, or Not To Keep: or Options In Between?

Hong Zhang
GSLIS, University of Illinois at Urbana-Champaign
501 E Daniel St.
Champaign, IL61820
1-217-244-2757
Hzhang1@illinois.edu

Michael Twidale
GSLIS, University of Illinois at Urbana-Champaign
501 E Daniel St.
Champaign, IL61820
1-217-265-0510
twidale@illinois.edu

ABSTRACT
With current systems, we are forced to make decisions either to keep or not keep; delete or not delete a file. Unfortunately our opinions about many information items do not easily fit into this binary worldview. As a part of an exploratory study of looking at file organization on personal computers, this paper describes how people deal with this difficulty on their computers. It implies that people need a facility for information items that falls between the categories of keep and not keep.

Categories and Subject Descriptors
H.1.2 [User/machine system]: Human information processing

General Terms
Management, Human Factors.

Keywords
Keywords are your own designated keywords.

1. INTRODUCTION
Many of us have more or less "pack rat" inclination, especially so on our computers. "I might need this some time", which is the typical reason for pack rat syndrome [3], applies to many of the files we saved on our computers. Falling costs and rising storage capacity in various forms and locations encourages this behavior of keeping files "just in case".

Unfortunately, keeping too much stuff on our computer has been seen as one of the main reasons for disorganization. Procrastination in filing and deleting is another main reason to cause clutter. Thus most personal productivity experts advocate being selective in keeping items, discarding items we no longer need, and preferably filing or discarding items right away. [4][5]

Current systems give us clear, simple binary choices. On first receiving a file from somewhere, we can either keep it or discard it. Later on we have the option of continuing to keep it, or of deleting it. But we are forced to make the decision. This can create a dilemma, or a distracting anxiety. Jones [1] describes the problem very well. On the one hand, not keeping information or not keeping information properly could cause access problem if/when it is needed. On the other hand, too much kept information could compete for attention that should be spent more on important items, and make finding other items more difficult. When structured organization information is important, keeping everything that might possibly be useful makes the act of organizing much more difficult. When seeing an information item, the decision of whether and how to keep it is not only "essential" but also "fundamentally difficult". The difficulty comes from the user's need to understand the information item, their own information organization space, and to anticipate future information needs. The decision of whether to delete older, previously filed files can be similarly difficult because again it rests on the anticipation of possible future needs.

As a part of an exploratory study of looking at file organization on personal computers, this paper describes how people deal with this difficulty on their computers, which implies that people need a facility for information choices that fall between keep and don't keep.

2. RELATED STUDIES
A number of studies have investigated keeping behaviors in paper world offices, especially how people use "piles" and "files" as two strategies for different needs. (e.g. [10][11])

Keeping behavior on personal computers has been studied quite thoroughly in Keeping Found Things Found (KFTF) project [1][2][7] and several other studies (e.g. [8][9]). A variety of keeping methods are observed in keeping Web information, and a set of functions are identified as the factors that influence the choice of method in different situations [7]. A special keeping behavior "clipping" which is defined as "intentionally saving portions of published material" is investigated in a study on how people save and use encountered information [8]. It has been recognized that with the increased ways, devices, and applications to keep information, and the increased number of information items we may keep in increased disk capacity, problems in personal information management such as information fragmentation and attention poverty could keep valuable information we saved from being used or even being noticed [9].

Among other findings, one of the key points about keeping that are identified in [2] is that "people don't always keep information with a specific purpose in mind", [2] which has resonated in this study.

In [2], William Jones defines "keeping" as:

Decisions made and actions taken to relate current information (information at hand or under consideration) to anticipated needs. Decisions can include (1) "ignore, this has no relevance to me," (2) "ignore, I can get back to this later", and (3) "keeping this in a special place or way so that I can be sure to use this information later."

This study takes a detailed look at the third keeping decision.

3. METHOD

6 Ph.D. students and 6 administrative staff in an academic environment were interviewed in front of their computers twice within 3 months. During the interview, they were asked to talk about several file/email folders and files. Interviews were audio recorded, and screenshots taken of selected folders. They were also asked to report via email their experience of information re-access difficulty during the 3 months.

4. RESULTS

4.1 Keep and Delete Behavior

In this study, when talking about what they do and do not keep, the Ph.D. participants talked a lot about whether or not to keep a item downloaded from the web, while administrative staff participants talked more about whether or not to delete an email message. The administrative staff talked about email as a major part of their work. They also deleted more in email than in their file system because of the constraint of space on the email server. As one administrative participant explained:

"…motivation for cleaning has more to do with space than anything else….I can't think of a time that I purposefully weeded specific folders because I haven't felt the need to because I have plenty of room to keep it all, so why not?"

Administrative participants sometimes have different keeping purposes such as keeping files for auditing. But it was found in both groups that they do not keep items categorized as "one time thing", or things "I don't think I'd need to come back", or "inconsequential things".

Another interesting keeping behavior observed in both groups of participants is a tendency to keep everything in a fast growing project. For example, one administrative staff had an email folder including 822 emails accumulated within a little over a month:

"it's just a lot of correspondence and I can probably delete some of it but it was just one of those things that when the project got started, it would be very complex and it was going to be a very fast project that had to get done right and I just felt like, just save it all and that way I won't have to worry about whether or not I lost something or misplaced some piece of correspondence that I needed to."

Not surprisingly, this participant subsequently encountered a difficulty with finding an email within this large folder. When asked about the problem, she believed that the keep everything strategy partly caused the problem:

"The fact that I am keeping every e-mail related to this project, instead of keeping selective e-mails, is part of the problem. Although it's good to have a record of all correspondence, it also creates a much larger volume of e-mails to sift through when looking for one in particular."

Interestingly, during a casual meeting after the two interviews, a Ph.D. student participant noted that he too had this "keep everything" behavior in files for a fast-growing project, without deleting or organizing.

4.2 Two types of special keeping

The study found that participants have clear sense of their own "main folder" or "home directory" which usually is under a root directory such as "My Documents", a hard drive, or a folder with the participant's name. They try to separate their files from those considered to belong to the computer system.

Within this context, we identified three types of "keeping" on computers, especially for Ph.D. student participants. Note that by saying "keep" the participants always meant "keep in main folders". In addition to information items that they want to keep in their main folders, there are items that they haven't decided whether or/and how to keep. Such items are made available for use right away by putting them on the desktop or in the root directory. These are referred to as "possibly useful" files in this paper. The third type are those files that participants do not intend to put into their main folders from the very start and only keep them on desktop or root directory for a while (referred as "to be deleted" in this paper).

It should be noted that in the collected data both "possibly useful" and "to be deleted" files were observed less frequently on administrative participants' file systems. Possible reasons for this are: a) the email system and the associated quota issue forced people to file messages to a folder, which changed the issue to be a "delete or not" problem as discussed in section 5; b) Ph.D. student participants' activities might be different from that of administrative participants in that there exist more vague needs of information items for future possible purposes.

Most of the "possibly useful" and "to be deleted" types of files were found being "dumped" either on the desktop or under the user's root directory (e.g. "My Documents"). This is partly because they are usually set as the default download or email attachment location. It is also because participants try to keep these files outside of their main folders in current systems.

4.2.1 "Possibly useful" items

The following examples illustrate people's descriptions of files that we are categorizing as "possibly useful":

"Papers are very good unless you go and read the section about it that says oh, this is …. Kind of figure out how to organize that stuff."

"…things I downloaded, for some reason or another I haven't got around to decide: a, if I need it, b, where it should be filed if I (need it)."

"These were here because at some point I want to get back to these to make sure there's nothing important. (you don't use them?) no, I might just delete most of them."

"…This I should figure out what it is. This (another one) I should figure out what it is…"

"Usually when I put something on my desktop, it's just because I want to take care of it quickly without thinking about when to put it."

"a lot of times the things on the desktop, I actually don't remember what they are, and I have to open them. …I have no idea what that is."

"(…the many files under My Documents?) I don't know if I want to file them or I haven't filed them yet. (Haven't used them for a while.) No. So it could be cleaned up."

The default download place can cause re-access difficulty. For example, one participant described his problems in trying to re-find an article he knew he had downloaded. The article had been downloaded to his desktop as the default place for downloads. Since the participant does not use the desktop (he was a Linux user), he went through a few searches before he finally found it. One Ph.D. student participant actually created a separate folder "webdownload" on the desktop for downloaded files, instead of leaving them mixed with other items on desktop:

"…when we didn't have any folder, when we downloaded something, it's on the desktop, and after some time it's very messy. So I told xx to create it and set it as default folder. Every download and attachment will go to here automatically."

Similarly, another participant packed these files up into a folder after a little deleting and filing at a cleanup time:

"…I just created this folder, that's called 'desktop articles' that's everything that was on the desktop."

An administrative participant also created a folder for email attachments:

"if I didn't know if I've saved, I just put in my attachment folder, so in later, I can go back and clean this out and put in various other folders."

Considering the amount of effort participants have to spend to clean up these files individually, it is not surprising that these "possibly useful" files are a major part of procrastination in organizing one's personal information. The above pack-up solution implies the need of a special kind of container for the files that are "possibly useful" but may only be checked rarely. On the Macintosh, the automatic download place utility folder plays a similar role in terms of a different place, but it is designed for system use and is separate from desktop use by users.

4.2.2 "To be deleted" items

The other type of special keeping is the "to be deleted" item. For example:

"…a lot of things that ended up on the desktop are really temporary, like I'm just looking at it for the time being. For instance, I have the current xxx conference schedule. I don't need to save that. So at some point, I just put it in the trash."

"…these are screenshots. ...because I need to show students how to do screenshots on Mac. Actually I can clean them up."

"…this actually was because I couldn't get it to print yesterday. I'm gonna get rid of it actually."

Similar to the "packing up" method for "possibly useful" stuff, another participant created a "temp" folder on the desktop:

"Usually after some time, if there are too many, I will delete it. …I know everything here is safe to delete."

Again, a separate container for these "to be deleted" files will make cleanup work much easier.

Example items of the "to be deleted" type:

- *"If it's something that I'm going to forget where it is anyway, I won't bother to download it."*
- *"I have the reading list so if there's anything I need to find, I can find it again, …it's easy enough to find electronic copies."*
- A conference schedule that *"usually goes to the desktop and I usually throw it out once the conference is over with."*
- print out, or use the *"online bigger network space"* instead of keeping many files on computer.
- rely on printed out paper copies, don't keep electronically.

The decision to keep or not is a judgment call at a given moment, and participants acknowledged that they could be wrong.

4.3 Delete

If we look at keeping as a decision made at the stage of receiving or retrieving an item, then "delete" is a decision of "not keeping" at a later stage, applied to files that have been already saved in main folders, when their usefulness has expired or their uselessness has become clear. Similar to the files in the previous section, there seems to be no systematic mechanism to clean up these files, except for triggers such as a space problem. For example:

"To me that's low priority. I'm not gonna mess with that unless I have spare time. …What I do is like today is slow day, I need to clean up those files because I got messages (for space quota) they want me to clean it up. …then I delete it."

"Probably these I need to clean up and delete. But I haven't had time and I don't care."

"…19 gigs is still a lot of space, but when it starts getting down to like 10 gigs or lower, then I'll start to worry and go clean."

Related to the "pack rat" inclination, several participants note being afraid of deleting things. For example:

"In the moment I thought it would be a good idea to keep it here. And then once things are there, I'm afraid to get rid of them. Whenever I've decided to save something, I'm afraid to get rid of it. I don't know, I might need it."

One administrative participant kept nearly all emails "even for one time use, I archive them. Only delete the spamming ones."

The decision to delete or not is another judgment call at a given moment and participants realize they "may (have) made mistakes to delete something that I don't mean to do".

"that has happened. I believe it has happened. …I wish I kept that, well but I didn't, and I just move one and figure out some other way to recall what it was …or re-create or whatever's the specific need at the time."

One administrative participant talked about the experience of trying to locate a deleted file, and finally "chose to redo it". Another administrative participant reported having the problem of

looking for a sent email after she deleted the sent folder because of quota problem and then had a two week vacation. Several participants keep trash for a while as a way of dealing with this problem:

"Some system, within 24 hours, it automatically dumps out the trash. That doesn't work well for me. I need a longer time to realize that an error was not created."

"…I rarely empty my trash. So a lot of times I went back to trash to see if I throw away something that I shouldn't have. …not too much, but I have definitely done that."

Several administrative participants had personal experience (and knew other colleagues with the same experience) of asking the help desk in the department to recover a file from the backup tape.

This implies that trash on a computer may need to have different levels. Files that were once useful, filed to the main folder and then get deleted are different from other useless or even spamming items, and can be packed and compressed in a way that people can still recover a file from it if needed, similar to the administrative participants' backup tape mechanism.

5. IMPLICATIONS & DISCUSSION

There are many files whose status seems somehow to fall between the binary decisions of to keep or not to keep. Allowing for a status between keeping and not keeping and accepting procrastination may help alleviate the filing or deleting difficulty that can be encountered in certain circumstances.

Providing containers for both the "possibly useful" and the "to be deleted" files will separate them from other files where clear binary decisions can be made. This may help decrease the clutter that competes for human attention with the more important items. It may also make organization easier by turning individual actions into batch dumping or batch packing actions at clean up time.

With the container for "possibly useful" stuff, for example, the items can look similar to what they are in current system so people can use them, except that they are in groups.

But at cleanup time, they can easily be packed up to have less visibility and less storage space. Search and even browse functions would then make retrieving items from it much easier.

Similar to other studies' findings (e.g. [6]), this study found that people sometimes do not save just because finding it later would be more difficult than finding it again on the Web. The large network becomes a part of personal information resource, as shown in the listed quotes in 4.2. The container for "possibly useful" items can be even extended to one for "the items I have seen before". One participant reported a re-access difficulty experience in email about trying to find something that "I know I read it somewhere, but can't remember what paper it's in". The container would be able to serve this need by doing a search.

Although keeping and organizing are related largely in both physical and digital world, this study proposes the possibility to "keep but not organize" certain types of information. Instead of fighting with human nature and limitation in front of vagueness and uncertainty, we might be able to deal with them in a more comfortable way.

6. REFERENCES

[1] Jones, William. 2004. Finders, keepers? The present and future perfect in support of personal information management. First Monday. DOI=http://firstmonday.org/issues/issue9_3/jones

[2] Jones, William. 2007. Keeping Found Things Found: The study and practice of personal information management. Morgan Kaufman Publishers.

[3] Warren, L.W. and Ostrom, J.C., 1988. Pack rats: world class savers. *Psychology Today* 22, pp. 58–62

[4] Barbara Etzel and Peter Thomas. 1996. Personal information management: tools and techniques for achieving professional effectiveness. New York University Press.

[5] http://www.timethoughts.com/timemanagement/effective-filing.htm

[6] Jones, William, Bruce, Harry, Dumais, Susan. 2001. Keeping Found Things Found on the Web.

[7] Jones, William and Dumais, Susan. 2004. Information behavior that keeps found things found. Information Research. 2004, Vol. 10, No. 1.

[8] Marshall, C. and Bly, Sara. 2005. Saving and using encountered information: implications for electronic periodicals. CHI2005.

[9] Marshall, C. and Jones, W. 2006. Keeping encountered information. Communications of ACM. 2006, Vol. 49, No. 1. p66-67.

[10] Malone, T.W. 1983. How do people organize their desks: implications for the design of office information systems. ACM Transactions on Office Information Systems. Vol. 1, No. 1, p99-112.

[11] Mander, R., Salmon, G. and Wong, Y. Y. 1992. A 'pile' metaphor for supporting casual organization of information. Proceedings of the ACM SIGCHI conference on Human factors in computing systems. P627-634.

Collaboration in an Open Data eScience: A Case Study of Sloan Digital Sky Survey

Jian Zhang
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104
001-215-895-2474
jz85@drexel.edu

Chaomei Chen
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104
001-215-895-2474
cc345@drexel.edu

ABSTRACT
Current science and technology has produced more and more publically accessible scientific data. However, little is known about how the open data trend impacts a scientific community, specifically in terms of its collaboration behaviors. This paper aims to enhance our understanding of the dynamics of scientific collaboration in the open data eScience environment via a case study of co-author networks of an active and highly cited open data project, called Sloan Digital Sky Survey. We visualized the co-authoring networks and measured their properties over time at three levels: author, institution, and country levels. We compared these measurements to a random network model and also compared results across the three levels. The study found that 1) the collaboration networks of the SDSS community transformed from random networks to small-world networks; 2) the number of author-level collaboration instances has not changed much over time, while the number of collaboration instances at the other two levels has increased over time; 3) pairwise institutional collaboration become common in recent years. The open data trend may have both positive and negative impacts on scientific collaboration.

Categories and Subject Descriptors
G.2.2. [**Graph Theory**]: Network problems

General Terms
Measurement, Theory, Verification.

Keywords
Open Data, Coauthor Network, Social Network Model, Small-world Network, Topological Analysis.

1. INTRODUCTION
Current science and technology has produced more and more publically accessible scientific data[8]. Many scientific projects primarily aim to collect scientific data, such as Sloan Digital Sky Survey (SDSS), Large Synoptic Survey Telescope (LSST)[1], and Ocean Observatory Initiative (OOI)[2]. These valuable scientific data are widely accessible through an eScience infrastructure to not only the targeted scientific communities, but other disciplines and the general public as well.

Impacts of this open data trend on scientific research in general and scientific collaboration in particular, has yet been widely studied, however. Openly accessible data may build up a common ground for scientists from different institutions, disciplines, and countries. Therefore it is possible to boost scientific collaboration. On the other hand, open data could lead to competitions for publishing the first discoveries, hence perhaps hindering collaboration. This paper aims to enhance our understanding of the dynamic patterns of scientific collaboration in the open data eScience environment, and characterize the impact of open data on science collaboration. As it is believed that collaboration can promote research activity, productivity, and impact[17], knowledge about these patterns may help future scientific projects, funding agencies, and scientists to foster and benefit from collaboration.

One way to define the existence of collaboration is through co-authoring relations found in scientific publications[9]. It has long been realized that co-authorship of scientific articles provides an informative unit of analysis for studying patterns of collaboration in scientific communities[25]. In this study, we adopted this perspective and investigated scientific collaboration in the open data eScience environment via a case study of co-authorship networks of Sloan Digital Sky Survey (SDSS), which is a highly cited open data project[8,28]. The SDSS survey is one of the largest digital sky surveys up date. It collects multiple types of data of stars, galaxies, quasars, and other astronomical objects in the universe. From 2000 to the date, the SDSS has produced 30TB astronomical data, and released these data to the scientific communities and the general public through SkyServer website[3]. These data have become a real gold mine for various scientific communities and good educational materials for the general public.

Conference'04, Month 1–2, 2004, City, State, Country.

[1] www.lsst.org
[2] www.oceanleadership.org
[3] skyserver.sdss.org

A comprehensive study of the impact of the open data trend on scientific collaboration would require comparisons between open data and non-open data projects or domains. This study mainly focuses on one particular open data project. We believe that as the first step to the study of the emerging open data phenomenon, painting a full spectrum of collaboration patterns in one project could lay down the foundation for future comparative studies in that the insights found in one project would help to generate and refine additional hypotheses and research questions for future studies.

The rest of the paper is organized as follows. Section two reviews the related work, and section three reports the methods and data used in this study. Section four presents the results of this study, while section five analyzes the most interesting results. Section six gives the conclusions of this study.

2. RELATED WORK

Co-authorship has long been used to study scientific collaboration. It can be traced back to the 1960s when Price and Beaver[27] used co-author relations to investigate social structures and influences in scientific communication networks. They concluded that the research front of a scientific domain is dominated by a small core of active researchers and a large weak transient population of their collaborators. In later years, Beaver and Rosen systematically explored co-authorship in a series of papers[4,5,6]. And various research communities use co-author relations to map research teams and collaboration structures. For instance, [21] mapped the research departments at two universities; and [26] focused on a chemical department's collaboration.

Collaboration based on co-authorship also could be aggregated at different levels of granularity, such as institutional, interdisciplinary, topical, or international collaboration, in which two entities are considered to be in collaboration if scientists from the two entities have co-authored one or more publications. Studies[15,20] of these aggregated levels of collaboration endeavored to understand the collaborative behavior across institutions, disciplines, and countries. Cummings and Kiesler found that multi-university collaboration created more problems than multidisciplinary collaboration[15]. A study of publications in high-energy physics found that although computer-mediated communication was believed to boost the intercollegiate and international collaboration, the percentage of papers having intercollegiate and international authorship increases evolutionary, rather than revolutionary[20].

Co-authorship relation can be easily transformed into a network structure, where nodes in the network represent authors, or other entities like institutions and countries, and edges represent co-authorship relations. A co-author network is one kind of graphs, various graph properties such as betweenness, bridge, centrality, clustering coefficient, degree, path length, and structure hole, could be measured to identify key nodes and edges. Exemplar studies include Chen's study of betweenness centrality of nodes to identify pivotal points in the evolution of scientific co-authors network[10], and Heinze and Bauer's study of structure holes in nano science and technology field to identify the brokerage role played by highly creative scientists in co-authorship networks.

In recent years, several social network models were applied to characterize scientific collaboration networks, such as the Erdos-Renyi model (random graphs[16]), the small-world model[30], and the scale-free model[1]. Social network models effectively describe topological characteristics across a wide range of large co-author networks in different disciplines such as biomedicine, high energy physics, astrophysics, mathematics, and computer science[22,24,25], neuroscience and mathematics[2], computer science[18], and condensed matter[9]. Newman[22,23] found that co-author networks all have a generic feature of a small-world network: a surprisingly short average distance (L) and a large clustering coefficient (C), much larger than the one expected from a random network with a same number of nodes and edges. Other studies confirmed this observation and found the value of L and C varied from discipline to discipline and from database to database.

While some social network model studies focused on a static overview of a collaboration network within a certain time frame, some studies looked at the dynamics of structural patterns of scientific collaboration networks over time[2,7,9,18]. Both Barabasi et al, and Huang used a "snowball sampling" approach to aggregate the publications at different time points (mainly in year). For example given a time point T between T_{start} and T_{end}, the publications used to construct a co-author network would be all the publications from T_{start} to the time point of T. And their studies found that while the number of authors and co-author relations kept increasing, the average distance and clustering coefficient kept decreasing in neuro science and mathematics database. Cardillo et al, directly divided publications in condense matter into one year time piece. They found the average distance slightly increased from 3.18 to 3.62 from 2000 to 2005, while the clustering coefficient is nearly constant around 0.71 throughout the six years.

A co-authorship network, to some extent, could be considered as a knowledge diffusion network in that conducting research and writing a paper is a knowledge exchanging and sharing process. Particularly, it could be a good strategy for developing institutions and countries to gain new knowledge via collaboration with advanced institutions and countries[29]. The structure of a small-world network is believed to be more efficient than that of a random network in terms of knowledge diffusion[14]. Morone and Taylor, however, found that the efficiency of knowledge diffusion in a small-world network depends on the initial "knowledge gap" among network members.

In summary, a co-author network can reveal structural patterns of scientific collaboration. Network topological analysis provides information for understanding the structure of collaboration networks as well as certain information for understanding knowledge diffusion. Our research aims to utilize these methods to reveal the dynamic patterns of scientific collaboration in the SDSS publications at different aggregated levels and improve our understanding of how the open data trend impacts scientific collaboration.

3. METHODS

Our method consists of three steps, including data acquisition and cleanup, co-author network generation, and network analysis.

3.1 Dataset

The SDSS literature dataset was collected from Thomson ISI's *Web of Science* (WoS). The data were retrieved with search terms: " 'SDSS' OR 'Sloan Digita*' " over a time span between 2001 and 2008. A total of 2252 records were retrieved. Since the WoS has a multidisciplinary coverage, the dataset may include some records that are not relevant to the SDSS project. The abbreviation SDSS has been used for terms other than the Sloan Digital Sky Survey, for example, Strategic Decision Supporting System. In these cases, we manually removed these irrelevant records by using functions in the WoS, such as analyzing the "Document Types" and "Sources Titles". We removed records identified as "Corrections," "Letters," "Editorial materials," and "Meeting abstract." We also removed records from 51 journals and conference proceedings that clearly have nothing to do with the SDSS, like Water Resource Management, Diabetologia, and Cancer. The final data collection includes 2138 bibliographic records of papers.

To better identify the entities, we retrieved the metadata of each record, including the authors, their affiliations, and the countries where these affiliations are located, and compared each pair of these entities to avoid inconsistency in the data. For example, some authors put the "Los Alamos Natl Lab**s**" for "Los Alamos Natl Lab", while some authors misspell "Apache Point Observ" with "Apach***u*** Point Observ" (mismatch is in bold and italic for readability).

The author name ambiguity is a long existing problem. Recently, some ID systems, such as ResearchID at Thomson Reuter and OpenID at OpenID Foundation, have been proposed to help solve this problem. Our dataset, however, does not contain this information. Therefore, we barely used the combination of full last name and first name initials as authors' identifier. In terms of institutions and countries, we applied a levenshtein distance measurement[19] to compare two strings of affiliations and countries, by which we unify the same institution and country that has different appearances in our dataset, and also correct some inconsistency caused by typos.

3.2 Generating Co-author Networks

Co-author networks and two aggregated collaboration networks were created in CiteSpace[11]. In order to have component separated views of these network, which make identification of the largest components easier, we regenerated these networks in Pajek[3]. The network files were converted into an edgelist format for network analysis in NetworkX toolkit[4].

3.3 Network Analysis

The collaboration network analysis focuses on topological analysis, which employs various statistical measures to characterize the topology of collaboration networks[2].

- Network size: we report the number of nodes and edges. The network size shows the size of research community of SDSS related study.
- Largest component (*LC*) size: A component is an isolated sub-network in a disconnected network. The largest component has the largest number of nodes among all components. We report the size of the *LC* with its number of nodes and edges in each network.
- Average distance (*L*): The average value of the shortest path length between any pair of nodes in a network. A shorter average distance means the collaboration between any pair of entities is closer. The average distance in a Erdos-Renyi model random network L_{rand} with the number of *N* nodes and *M* edges is obtained using the following formula ([13] p. 144):
  $L_{rand} = \ln(N)/\ln(2M/N)$
- Clustering coefficient (*C*): A network's clustering coefficient is the average clustering coefficient (*c*) over all the nodes, which is calculated as the ratio of the number of edges between the node's direct neighbors to the number of possible edges between the node's direct neighbors.

  c= (the number of edges between the neighbors)/(the possible number of edges between the neighbors)

  The clustering coefficient in a Erdos-Renyi model random network C_{rand} with the number of *N* nodes and *M* edges is obtained with the following formula[16]:

  $C_{rand} = 2M/N(N-1)$

From these topological measures, one can characterize and compare collaboration networks. Along with a time line, the dynamic patterns of collaboration can also be observed.

4. RESULTS

Collaboration networks and their topological measurements are presented in this section in the order of author-level collaboration, institutional collaboration and international collaboration. Then we report the comparison of the topological measurements across the three levels along with the dynamic pattern.

4.1 Collaboration Networks and Topological Measurements

4.1.1 Author level collaboration in SDSS

Figure 1 shows snapshots of the eight-year author level collaboration networks. Nodes are in red and edges are in grey scale. The eight year author collaboration networks are all

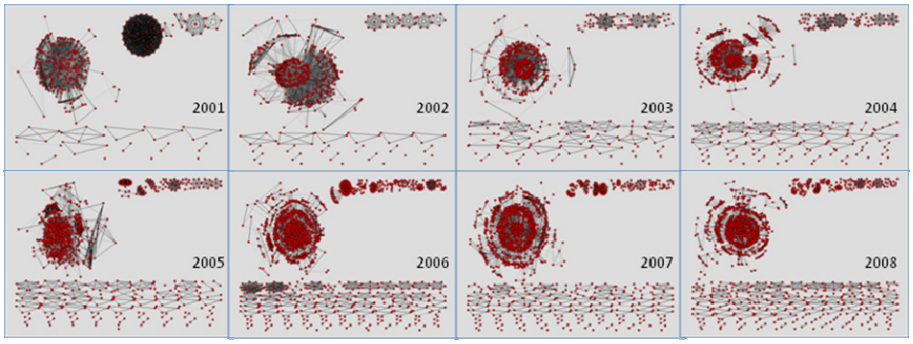


Figure 1. Author collaboration networks in separated component view (2001-2008).

[4] networkx.lanl.gov

Table 1. Topological measurements of author collaboration networks

Year	*N*	*E*	*C*	C_{rand}	*L*	L_{rand}	*LC-N* (%)	*LC-E* (%)
2001	324	7279	**0.623**	**0.139**	1.603	1.519	231(71.3)	6583(90.4)
2002	507	24407	**0.828**	**0.190**	1.777	1.364	425(83.8)	24264(99.4)
2003	616	21978	**0.861**	**0.116**	1.891	1.505	458(74.4)	21641(98.5)
2004	846	19702	0.779	0.055	2.332	1.755	641(75.8)	19210(97.5)
2005	1030	20088	0.817	0.038	2.629	1.894	743(72.1)	19394(96.5)
2006	1728	26945	0.694	0.018	2.624	2.167	1198(69.3)	24397(90.5)
2007	1533	23780	0.813	0.020	2.7	2.136	1125(73.4)	22649(95.2)
2008	1692	29028	0.766	0.020	2.81	2.103	1216(71.9)	27924(96.2)

N: the number of nodes; *E*: the number of edges; *C*: clustering coefficient; C_{rand}: clustering coefficient of a random network of the same size; *L*: average distance; L_{rand}: average distance of the random network; *LC-N* (%): the number of nodes in the largest component and the percentage to the total number of nodes; *LC-E*(%): the number of edges in the largest component and the percentage to the total number of edges.

dominated by one giant cluster. Besides the largest cluster, some authors formed relatively small clusters, and few authors worked alone, hence becoming single nodes without edges connected to the other nodes.

Table 1 lists the topological measurements of these networks. The number of nodes in the SDSS collaboration network increased almost linearly throughout the eight years, while the number of edges, the co-author collaboration, has two drops at year of 2004 and 2007, but remains relatively constant.

In the early years (2001 to 2003), the values of clustering coefficient in author collaboration networks are relatively close to the random network at the same order of magnitude (highlighted in bold font), while in the later years the differences become larger at different order of magnitudes. The average distance of author collaboration network increased almost linearly in the eight years. The average distance is larger than the average distance of the random network, but still in the same order of magnitude.

As depicted in Figure 1, a large number of nodes formed the largest component (around 70% to 80% of the total number of nodes), and these nodes forms the majority of collaborative relations (more than 90% of the total number of edges).

4.1.2 Institutional collaboration in SDSS

Figure 2 shows snapshots of institutional collaboration networks over the eight years. The networks in Figure 2 also have a large dominant cluster with several much smaller clusters. The evolution of the institutional collaboration is analogous to a tree-like process, in which a few nodes in the middle of the largest cluster form a core; while the other nodes in this cluster expend from the core with few trunks, but there are very few interconnections among these peripheral nodes.

Table 2 lists the topological measurements of institutional collaboration networks in SDSS. The number of nodes and edges were increasing in the eight years except for a drop at 2007. Compared to a random network, the gaps in terms of the values of clustering coefficient were larger in 2001, then close in 2002 to 2004 in the same order of magnitude (highlighted in bold font), and became larger in the rest of the years. The values of clustering coefficient in institution networks were decreasing over time. The average distance in the institutional collaboration is smaller than the average distance in random networks for all years, and shows an increasing tendency.

Table 2. Topological measurements of institute collaboration networks

Year	*N*	*E*	*C*	C_{rand}	*L*	L_{rand}	*LC-N* (%)	*LC-E* (%)
2001	46	41	0.22	0.040	2.402	6.623	28(60.9)	33(80.5)
2002	74	59	**0.062**	**0.022**	2.234	9.224	36(48.6)	36(61.0)
2003	104	110	**0.079**	**0.021**	3.137	6.199	72(69.2)	92(83.6)
2004	153	139	**0.073**	**0.012**	2.914	8.424	87(56.9)	109(78.4)
2005	190	163	0.074	0.009	2.807	9.719	96(50.5)	118(72.4)
2006	262	265	0.077	0.008	3.706	7.904	166(63.4)	219(82.6)
2007	260	252	0.032	0.007	3.383	8.401	156(60.0)	205(81.3)
2008	281	267	0.036	0.007	3.889	8.782	184(65.5)	228(85.4)

The percentage of nodes and edges in the largest cluster to the total number of nodes and edges fluctuated in the eight years, showing no clear trends.

4.1.3 International collaboration in SDSS

Figure 3 shows snapshots of the eight years international collaboration networks. There is also a major cluster in all networks. Visual observation reveals that the major component is a tree in 2001, and several core nodes formed the center of the networks in 2002 and 2003, and then the density of the networks increased in the more recent years. There are only a few nodes isolated from the major component, and the number of these nodes is less than ten.

Table 3 lists the topological measurements of the international collaboration networks. Similar to the institutional collaboration networks, the total number of nodes and edges in international collaboration networks has an increasing tendency except a drop at 2007.The values of clustering coefficient in country collaboration networks show the similar tendency as institution collaboration networks, which is close to random networks in the early years (highlighted in bold font) and became much larger in the later years. The values of average distance in international collaboration networks are smaller than that of random networks, and fluctuated in the eight years, showing no clear trends.

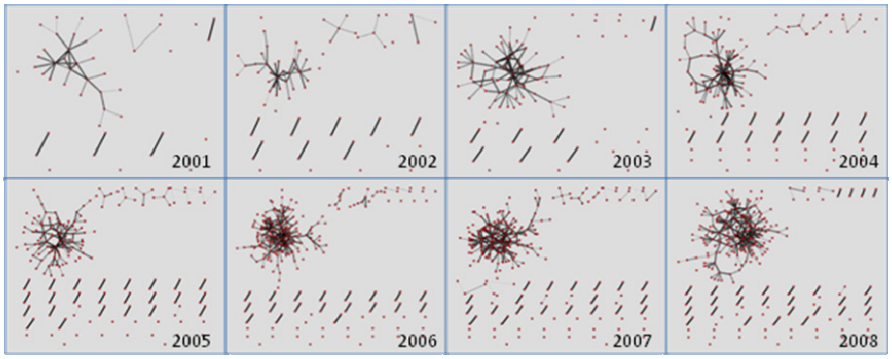


Figure 2. Institute collaboration networks in separated component view (2001-2008).

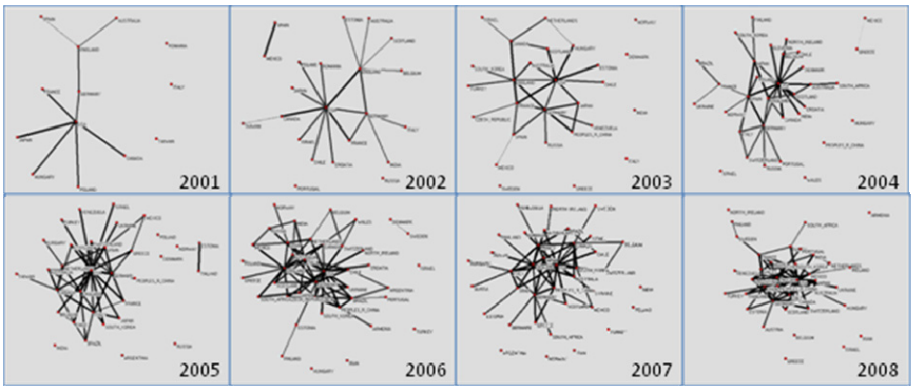


Figure 3. International collaboration networks in separated component view (2001-

Table 3. Topological measurements of international collaboration networks

Year	*N*	*E*	*C*	C_{rand}	*L*	L_{rand}	*LC-N* (%)	*LC-E* (%)
2001	13	11	0	0.141	2.035	6.623	11(84.6)	10(90.9)
2002	22	20	**0.083**	**0.087**	2.061	9.224	18(84.6)	19(95.0)
2003	27	30	**0.083**	**0.085**	2.061	6.199	18(81.8)	29(96.7)
2004	33	38	0.179	0.072	2.312	8.424	27(81.8)	37(97.4)
2005	33	52	0.23	0.098	1.746	9.719	25(75.8)	49(94.2)
2006	40	70	0.277	0.090	2.236	7.904	34(85.0)	69(98.6)
2007	38	58	0.281	0.083	2.31	8.401	32(84.2)	58(100.0)
2008	39	73	0.258	0.099	2.469	8.782	35(89.7)	73(100.0)

These largest components in international collaboration networks contain 80 to 90 percent of nodes and nearly all edges (95% to 100%).

4.2 Comparison of topological measurements across the three levels

This section shows the comparison of the topological measurements across the three level collaboration networks.

Figure 4 depicts the network sizes at the three level collaboration networks over the eight years. Because the author level collaboration networks have a large number of nodes and edges than corresponding networks at the other two aggregated levels, we use logarithmic scale in the y axis. Figure 4 shows that the number of nodes at all three level collaboration networks increased in a similar pattern. The increasing number of nodes means the research community, including scientists, institutions and countries, of SDSS grew continuously. Surprisingly, the

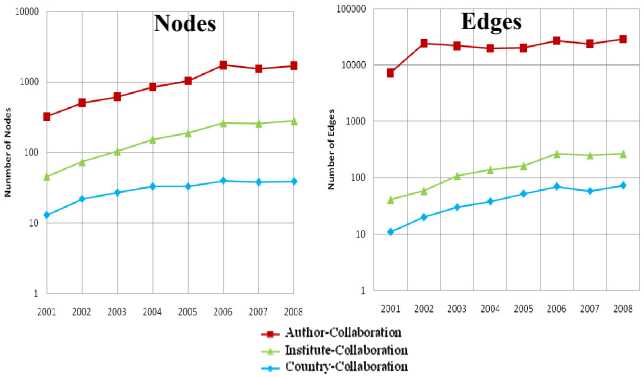


Figure 4. Dynamic trends of network sizes in SDSS.

number of edges suggests a different scenario. While the numbers of edges in institutional and international collaboration networks increased with the similar trend as nodes did, the number of edges in author collaboration networks is nearly constant in the eight years. Therefore the average degree (the ratio of two times of the number of edges to the number of nodes) in author collaboration networks decrease, which means the average number of collaborators of a scientist decreased in SDSS collaboration.

Figure 5 shows the comparison of clustering coefficient across the three levels over time. Author collaboration networks have larger values of clustering coefficient than corresponding networks at institutional and international levels, which is expected since the authors from the same institution tend to collaborate frequently than authors from other institutions and countries. International collaboration networks have larger clustering coefficient values than institutional collaboration. Because this study ignored the weight information of edges, a single publication co-authored between two countries' scientists will establish a collaboration link and be treated equivalently as links that represent many instances of collaboration. In this case, the collaboration in country level is expected to denser than institutional collaboration.

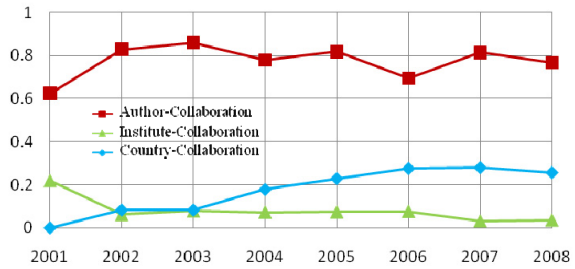


Figure 5. Dynamic trends of clustering coefficient in SDSS.

In terms of dynamic trends, the institutional collaboration networks show an interesting decrease, which means pairwise collaboration relations became popular, and two institutions that collaborated with the same third party institution is less likely to collaborate with each other. International collaboration has reverse tendency, as the values of clustering coefficient in country increased in the majority of years, two countries that collaborated with one common country become more likely to

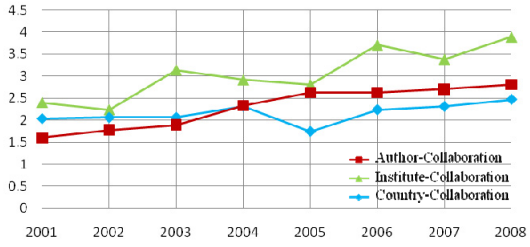


Figure 6. Dynamic trends of average distance in SDSS.

collaborate together.

Figure 6 shows the trends of the values of average distance in three collaboration networks. The institutional collaboration networks have the longest average distance than the other two levels, and kept increasing while fluctuated several times. The longer average distance values mean that in general to establish a collaboration relation between two institutions is harder than find another person or country given the long intermediate institutions that needed to pass by. The average distance in author collaboration networks has an ascent tendency, from 1.5 to nearly 3, which means as the size of the SDSS community increased, finding another collaborator needs to pass more persons. The international collaboration has nearly the same average distance over the eight years, and is the smallest one.

The results in Figure 5 and 6 together show an interesting pattern. While the clustering coefficient values in institutional collaboration networks kept decreasing, their average distance went up, suggesting that the institutional collaboration networks become sparser and more tree-like shape. This analysis result is consistent with direct observation in Figure 2.

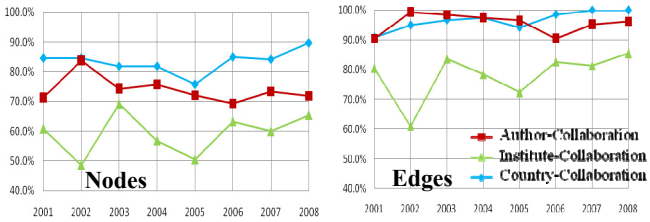


Figure 7. Dynamic trends of proportions of the largest component to the entire network in SDSS.

Figure 7 shows the proportions of the number of nodes and edges in the largest components to the total number of nodes and edges in the network. In the three-level collaboration networks, Figure 7 shows no clear trends. Compared to other two levels, institutional collaboration has the lowest percentage of the number of nodes and edges in the largest components in all eight years. Nearly 40 percent of nodes and 20 percent of edges are outside the largest components. In author and international collaboration networks, the nodes isolated from the largest component are more likely to work along given the fact that they only count for less than 5% of the edges in the whole network. Figure 7 shows that in SDSS–related studies 60 percent of institutions dominate the majority collaboration relations, but still many institutions (around 40%) can carry on their own research by using SDSS data.

5. DISCUSSIONS

This section focused on the implications of the results in above section. Due to the paper length limit, we only highlight the most interesting and unexpected patterns derived from the results, including the random to small-world evolution patterns in all three level collaboration; the relatively constant number of author level collaboration instances versus the increasing number of collaboration instances at the other two levels; and the tree-like network evolution process in institutional collaboration networks. We also discuss some limitations of this study.

First, in all three level collaboration networks, the clustering coefficient values all show a random to small-world evolving pattern. In the early years, the collaboration networks have clustering coefficient values in the same order of magnitude to the random networks of the same size. In author level networks, the average distance in early years are also very close to the corresponding values in random networks. In later years, the collaboration networks all have larger clustering coefficients than random networks do. In the institutional and international collaboration, the average distance in collaboration networks is smaller than random networks, but still in the same order of magnitude.

The random to small-world evolution pattern may represent the actual process of the SDSS project. In the early year, a few scientists who were active members of the SDSS research community may randomly choose their collaborators to start research collaboration, and then when collaboration relationship grew up, the network became more like a small world network, where acquaintances to a common acquaintance become acquainted with each other. Examples may be shown as in [12], scientists tend to collaborate with Chilean institutions who have astronomical facilities that can help verify their discoveries from the SDSS data. The open data environment may facilitate the process since they have a common ground — the same SDSS data.

Second, even though the number of collaboration relations in author level varied, it shows a relatively constant tendency, while at the other two levels, the number of collaboration relations increased over time. We may posit a plausible explanation for this phenomenon. In astronomy community, the phase of postdoctoral training is common, hence generating a circulation of young astronomers between various schools and groups. When postdocs moved around different institutions and countries every few years, they may still tightly collaborate with the same group of astronomers, but in effect increase the chance of institutional and international collaboration relationship. In this process, the SDSS data, which is widely accessible in different locations, may help to maintain the tight collaboration relationship among the postdocs and their collaborators. However, to answer whether open data is the only reason that can explain this phenomenon, future studies are needed with carefully designed comparisons of collaboration networks between open data and non-open data environments.

Third, the direct observation of network images and the topological measurements confirm the tree-like evolving process of the institutional collaboration networks. The decreasing values of clustering coefficient and increasing values of average distance in institutional collaboration networks suggested that in SDSS research the possibility of establishing a new collaboration relation between two institutions that have a common collaborator become lower and lower.

Why is the pairwise collaboration in institutional collaboration very common in the recent years? Does the open data trend have impacts on this phenomenon? On the one hand, according to Cumming and Kiesler[15], multi-university collaboration is hard, requiring good coordination skills and supportive mechanisms. In our results, the pairwise collaboration could be a better tradeoff of innovation opportunities versus coordination costs than collaboration involved with three or more institutions. On the other hand, the publically accessible SDSS data may even hinder the institutional collaboration. When data is widely available, competing for the first publication of a discovery could prevent

different institutions from some collaborative engagements, like sharing methods and results, which are much likely and easier to take place within one institution.

Some limitations of this study possibly constrain the generality of the results.

First, this study describes the dynamic patterns in one open data eScience project without comparing to other open data projects, or non-open data studies. Hence explanations raised in this section could only be considered as exploratory impacts of open data on science collaboration. The discovered patterns, however, could lay down the foundation for further comparison studies.

Second, our study ignores the weight information in collaboration networks. All edges are considered to be equal no matter how many publications were co-authored by two nodes. This may bias the results, especially in higher aggregated level like country level. Further research could achieve more accurate results if it can take the weight information into considerations.

6. CONCLUSIONS

In this study we have investigated the dynamic collaboration patterns in an open data eScience environment via a case study of co-authorship relation in the SDSS publications. We studied the co-authorship collaboration networks at three levels, namely the author level, institutional level, and country level. By visualizing these collaborative networks and measuring their topological properties over time, we reached the following conclusions.

1) The collaboration networks in the SDSS community experienced an evolution from a random network to a small-world network. But the small world properties varied across the three different collaboration levels. The institutional collaboration has a much less like small-world ingredient, a tree-like evolving process.

2) In SDSS the number of collaboration relations at the author level is relatively constant, while at the institutional and country levels, the numbers were increasing. The open data trend could help to explain this observation, but future studies are needed to compare our results to other open data projects and non-open data projects.

3) Pairwise institutional collaboration became common in recent years in the SDSS community. The open access data may hinder the collaboration among multiple institutions.

To our humble knowledge, this is the first study focused on the collaboration patterns in an open data eScience environment. The methodology framework developed in this study can be used in other open data or non-open data domains, such as the iSchool community. More studies are eagerly needed to better understand the new open data trend in science and its impacts on science.

7. ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No. IIS-0612129. Thomson Reuters provides the bibliographic data for the analysis. We also thank the anonymous reviewers for their constructive insights to the early draft. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

8. REFERENCES

[1] Barabasi, A.-l., and Albert, R. 1999. Emergence of scaling in random networks. Science 286, 509-512.

[2] Barabasi, A.-l., Jeong, H., Neda, Z., Reavasz, E., Schubert, A., and Vicsek, T. 2002. Evolution of the social network of scientific collaboration. Physica A 311, 590-614.

[3] Batagelj, V. 2001. Pajek - program for large networks analysis and visualization. in Dagstuhl seminar Link Analysis and Visualization, Dagstuhl.

[4] Beaver, D.d.B., and Rosen, R. 1978. Studies in scientific collaboration. Part I. The professional origins of scientific co-authorship. Scientometrics 1, 65-84.

[5] Beaver, D.d.B., and Rosen, R. 1979. Studies in scientific collaboration. Part II. Scientific co-authorship, research productivity and visibility in the French scientific elite. Scientometrics 1, 133-149.

[6] Beaver, D.d.B., and Rosen, R. 1979. Studies in scientific collaboration. Part III. Professionalization and the natural hisotry of modern scientific co-authorship. Scientometrics 1, 231-245.

[7] Bettencourt, L.M.A., Kaiser, D.I., and Kaur, J. 2009. Scientific discovery and topological transitionsin collaboration networks. Journal of Informetrics 3, 210-221.

[8] Borgman, C.L. 2007. Scholarship in the Digital Age: Information, Infrastructure, and the Internet, The MIT Press, Cambridge, Massachusetts.

[9] Cardillo, A., Scellato, S., and Latora, V. 2006. A topological analysis of scientific coauthorship networks. Physica A 372, 333-339.

[10] Chen, C. 2005. The centrality of pivotal points in the evolution of scientific networks. in the 10th International Conference on Intelligent User Interfaces (IUI 2005), ACM Press, San Diego, CA USA.

[11] Chen, C. 2006. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. Journal of the American Society for Information Science and Technology 57, 359-377.

[12] Chen, C., Zhang, J., and Vogeley, M.S. 2009. Mapping the Global Impact of Sloan Digital Sky Survey. IEEE Intelligent Systems 24, 74-77.

[13] Chung, F., and Lu, L. 2006. Complex Graphs and Networks, California State University, San Marcos, CA USA.

[14] Cowan, R., and Jonard, N. 2004. Network structure and the diffusion of knowledge. Journal of Economics Dynamics & Control 28, 1557-1575.

[15] Cummings, J.N., and Kiesler, S. 2005. Collaborative Research Across Disciplinary and Organizational Boundaries. Social Studies of Science 35, 703-722.

[16] Erdos, P., and Renyi, A. 1959. On random graphs. Publicationes Mathematicae 6, 290-297.

[17] Glanzel, W., and Schubert, A. 2004. Analyzing scientific networks through co-authorship. in Handbook of Quantitative Science and Technology Research (Moed, H., Ed., Springer, Netherland.

[18] Huang, J., Zhuang, Z., Li, J., and Giles, L.G. 2008. Collaboration over time: characterizing and modeling network evolution. in First ACM International Conference

on Web Search and Data Mining (WSDM 2008), ACM Press, Palo Alto, CA USA.
[19] Levenshtein, V.I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10, 707-710.
[20] Lorigo, L., and Pellacini, F. 2007. Frequency and structure of long distance scholarly collaborations in a physics community. Journal of the American Society for Information Science and Technology 58, 1497-1502.
[21] Mahlck, P., and Persson, O. 2000. Socio-bibliometrics mapping of intra-department networks. Scientometrics 49, 81-91.
[22] Newman, M. 2001. Scientific collaboration network: I. Network structure and fundamental results. Physica Review E 64, 016131-1-8.
[23] Newman, M. 2001. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. Physica Review E 64, 016132-1-7.
[24] Newman, M. 2001. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences 98, 404-409.
[25] Newman, M. 2004. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences 101, 5200-5205.
[26] Peter, H.P.F., and Van Raan, A.F.J. 1991. Structuring scientific activities by co-author analysis: An expercise on a university faculty level. Scientometrics 20, 235-255.
[27] Price, D.d.S., and Beaver, D.d.B. 1966. Collaboration in an invisible college. American Psychologist 21, 1011-1018.
[28] Singh, V., Gray, J., Thakar, A., Szalay, A.S., Raddick, J., Bill, B., Lebedeva, S., and Yanny, B. 2006. SkyServer Traffic Report-The First Five Years. in Microsoft Technical Report MSR TR-2006-190, Microsoft.
[29] Wagner, C.S. 2008. The New Invisible College: Science for Development, Brooklings Institute Press, Washington, D.C.
[30] Watts, D., and Strogatz, S. 1998. Collective dynamics of "small-world" networks. Nature 393, 409-410.

Roundtables

iConference 2010

FEBRUARY 3-6 • UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

The Role of iSchools in Shaping the Future of Health Informatics

Kelly E. Caine
Indiana University
caine@indiana.edu

Kay Connelly
Indiana University
connelly@indiana.edu

Barb Hayes
Indiana University-Purdue University at Indianapolis
bmhayes@iupui.edu

Julie A. Kientz
University of Washington
jkientz@u.washington.edu

INTRODUCTION

In remarks to the National Academy of Sciences earlier this year President Obama discussed many of the proposed benefits of stimulating growth in health and medical informatics. He proposed that computerizing medical records would reduce duplication and waste, and would help to prevent errors that cost dollars and lives. He also noted that "records hold the potential of offering patients the chance to be more active participants in prevention and treatment" [1].

For this hope to become a reality, people need to be able to understand and use their own health information. Consumers of health care are increasingly being asked to take more responsibility for their health and pay a larger share of their health care bills. New digital tools are giving patients access to their own information and helping them interpret and act on that information. Nevertheless, there are many challenges in creating health information useful to consumers in understandable, usable, and actionable ways. Fortunately, many researchers in iSchools have already begun to address these challenges.

In this roundtable discussion we will explore how iSchools can shape the future of Health Informatics research and practice. We will introduce many of the modern challenges in Health Informatics including: how practitioners and consumers may collect, transform, synthesize, analyze and act on health related information and discuss how iSchools are uniquely positioned to address these interdisciplinary challenges. In addition, we will consider how the International Medical Informatics Association's [2] recommendations on education in health and medical informatics fit with the structure of iSchools.

This session will also focus on the ways iSchools can best prepare students for careers as health informaticians. We will discuss differences between medical informatics (see [3] for a discussion of medical informatics within iSchools) and health informatics and also explore differences between informatics classes taught in medical and nursing programs and informatics classes taught within iSchools.

SAMPLE QUESTIONS

Roundtable leaders will use several key questions to guide the discussion including:

- How can we take advantage of the strengths of iSchools to create the best learning environment for future health informaticians?
- How can we measure the success of Health Informatics programs in iSchools?
- How do existing iSchool courses fit into a new health informatics curriculum?
- What are the core competencies of a health informatician?
- What are the unique strengths of getting an education in health informatics from an iSchool?
- What are the impacts of housing health informatics programs within iSchools rather than medical and/or nursing programs?
- What are the unique challenges of teaching health informaticians within iSchools?

CONCLUSION

At the conclusion of the session, participants will have a broader understanding of the unique challenges and opportunities of Health Informatics education within iSchools. We expect that participants will bring a wide variety of knowledge and experience to the discussion and that by the end of the discussion everyone will have gained information on existing resources and have a better grasp on gaps which must be filled to deliver effective academic offerings and foster successful research outcomes in this arena.

DISCUSSION LEADERS

Kelly Caine is a Research Fellow in the School of Informatics and Computer Science at Indiana University. Dr. Caine's research interests are in health informatics, aging, privacy, and HCI. She is currently a member of the ETHOS group at Indiana University where she is collaborating on research investigating ethical technologies to support independent, healthy, at-home living for older adults. Dr. Caine received her Ph.D. in Psychology from the Georgia Institute of Technology in 2009.

Kay Connelly is an Associate Professor in the School of Informatics at Indiana University. Her research interests are in the intersection of mobile and pervasive computing and healthcare. In particular, she is interested in issues that influence user acceptance of health technologies, such as privacy, integration into one's lifestyle, convenience, and utility. Dr. Connelly works with a variety of patient groups, including very sick populations who need help in managing their disease, healthy populations interested in preventative care, and senior citizens looking to remain in their homes for as long as possible. Dr. Connelly received a BS in Computer Science and Mathematics from Indiana University (1995), and an MS (1999) and Ph.D. (2003) in Computer Science from the University of Illinois.

Barb Hayes is Associate Dean for Administration and Planning at the School of Informatics at Indiana University-Purdue University at Indianapolis. Hayes teaches social and organizational informatics and had a twenty-five year career in hospital-based healthcare before joining the faculty at IUPUI. She currently works on a number of health and life science strategic planning initiatives for IU and the Indiana business community.

Julie Kientz is an Assistant Professor in The Information School and the department of Human Centered Design & Engineering and at the University of Washington. She directs the Computing for Healthy Living and Learning Lab and is active in the dub Group alliance. Her research interests are in the areas of HCI, Ubicomp, and CSCW. In particular, her work focuses on how data collection and reflection can be made easier, more efficient, and more fun in health and educational settings. Dr. Kientz received her Ph.D. in Computer Science from the Georgia Institute of Technology in 2008.

REFERENCES

[1] Obama, B. 2009. Remarks by the President at the National Academy of Sciences Annual Meeting. Available online: http://www.whitehouse.gov/the_press_office/Remarks-by-the-President-at-the-National-Academy-of-Sciences-Annual-Meeting/

[2] International Medical Informatics Association. 2000. Recommendations of the International Medical Informatics Association (IMIA) on education in health and medical informatics. Methods of Information in Medicine. 39 (3); 267–77.

[3] Detlefsen, E., Hersh, W., Schardt, C. M., & Wildemuth, B. M. 2009. Alternative Approaches to Educating Medical Informationists.

Measuring the IMPACT of Early Childhood Information Literacy Programs on Children

Eliza T. Dresang
University of Washington iSchool
PO Box 352840
Seattle, WA 98195-2840
01-206-542-0169
edresang@u.washington.edu

Kathleen Burnett
Florida State University, SLIS
PO Box 3062100
Tallahassee, Fl 32306-2100
01-850-644-8124
kburnett@fsu.edu

Janet Capps
Florida State University, SLIS
PO Box 3062100
Tallahassee, FL 32306-2100
01-850-644-5775
Capps.janet@gmail.com

ABSTRACT

This roundtable discussion proposal supports the IMPACT theme of the conference as well as the conference area of interest in "Information behavior: theoretical, empirical and methodological advances in everyday life settings. . ..information literacy." It complements the new research stream for papers developed by the Associate Deans for Research of the iSchools, "Measuring Research Impact." And it brings focus on a user group, very young children, who are often overlooked in information research
The proposal for this discussion stems from an initiative in the state of Washington, which appears to have an exemplary and unique program for providing early learning, particularly early information literacy programs, in which libraries, public and school, play a leadership role. To find out whether what appears to be the case is indeed the case, the University of Washington Information School, the Washington Early Learning Public Library Partnership (ELPLP) consisting of 25 urban, suburban, and rural library systems, the Washington Foundation for Early Learning (FEL), a non-profit organization supporting early childhood development, and the Florida State University College of Communication and Information, have joined forces. The ELPLP, FEL, as well as a cabinet-level Department of Early Learning, appear to be unique to Washington.

In September, the four partner institutions, with the University of Washington as the lead, received a one-year National Leadership Planning Grant from the Institute for Museum and Library Services to develop a plan to assess or measure the impact on children of the early childhood literacy programs in which Washington libraries are involved. This partnership is firmly committed to developing a national model that includes an assessment of the impact (rather than the inputs) of library-related early learning programs on children
The researcher submitting this proposal developed, with other colleagues, a method for assessing impact of technology use in libraries for older children in the early part of this decade (see Dresang et al. Dynamic Youth Services through Outcome Based Planning and Evaluation). Although certain components of that model might apply to early learners, in general it is not developmentally appropriate for this age group.

Aspects of this round table discussion that will interest various iSchools Conference participants include

--reflection on the information needs of a user group that does not appear often in iSchool research, preliterate children and how information literacy is provided to these young users

-- proposed research methods, some of which are uncommon in the field of information, e.g., Bayseian networks, a statistical method involving probability of certain observable behaviors in relation to achieving desired learning outcomes

-- a tool that the doctoral candidate (Capps) is developing to measure core knowledge of early literacy providers (an input that will surely affect the outcome)

-- insight into a unique and very strong collaborative community partnership with library leadership that focuses on research results as a primary goal

-- the opportunity to contribute ideas to a multi-year information-related project that is under development.

Categories and Subject Descriptors

A. General Literature

General Terms

Measurement, Performance, Experimentation,

Keywords

Information seeking and use, literacy, research, methods

Native Systems of Knowledge: Indigenous Methodologies in Information Science

Miranda Belarde-Lewis, Marisa E. Duarte, Ally Krebs
The Information School
University of Washington
Seattle, WA 98195
{mhbl, meduarte, krebsa}@uw.edu

1. INTRODUCTION

Study of Native systems of knowledge involves examining the institutions, community practices, philosophies and policies around knowledge, information, and technology that support Indigenous and tribal sovereignty. For Indigenous peoples, the realities of colonialism reverberate in the management, representation, and design of information systems and services in Indigenous and tribal communities. Questions of who manages, owns, controls, and accesses Indigenous information systems and knowledge are crucial in understanding the relationship between economic and educational conditions in Indigenous communities, and global and local policies around information systems, intellectual property, and sovereignty. Anticolonial and decolonizing research strategies offer one way to leverage Indigenous knowledges to support autonomy and sovereignty.

In this roundtable, we will discuss how colonialism manifests in research on information with Indigenous communities, and strategies for designing and conducting anticolonial and decolonizing research.

The panel will present the following:

1) Three examples of research in information science utilizing Indigenous methodologies.
2) Relevant theoretical frameworks.
3) A discussion on how Indigenous methodologies grounds the study of information in the context of globalization.

2. PANELIST TOPICS

Through discussion of three different studies of information in the Indigenous context--data visualization, information as ideology, and institutional information practices—the panelists can speak to broader issues informing Native systems of knowledge, development of an Indigenous consciousness, and how government relationships shape information systems and services for tribal peoples.

2.1 Indigenous Information Visualization

Information visualization is a process that transforms data, information and knowledge into a form that relies on the human visual system to perceive its embedded information. [1] The transformation of tribally specific knowledge into visual form has been happening in Indigenous and tribal communities since time immemorial, yet has largely been erroneously perceived and has been limited in its categorization as "art." The visual literacy required to read the embedded codes of Native "art" present two major questions: 1) how does being literate in these visual means of communication, tribal history and knowledge help to shape, reinforce and assert a specific tribal identity, and 2) how does the storing, teaching, and access to the visual representations of a specific cultural group influence the narratives, stereotypes and literature about that group? [2] Broader implications of sovereignty, intellectual property, and traditional knowledge are encompassed in the study of Indigenous information visualization, especially when the perspective is from the community that is being visually represented in a variety of contexts.

2.2 Information in the Indigenous Consciousness

Scholars are analyzing the power dynamics inherent in information and knowledge structures, and research and technology development in Indigenous communities as a manifestation of imperialism. [3] This is evident in information and knowledge structures that objectify, decontextualize, or misrepresent the political reality of Indigenous peoples. [4] As a result of common experiences among different Indigenous communities dealing with non-Indigenous research teams and agencies implementing information policies and technologies, an Indigenous consciousness is being developed that sharply criticizes organizations with neo-liberal agendas. [5] [6] A core body of work is needed to address problems of information from the perspective of a critical Indigenous consciousness such that knowledge gained from research can enable Indigenous sovereignty and autonomy while also contributing to the broader discussion about the function, ethics, and methods of research and policy-making around information and technology in the globalized world.

2.3 Indigenous Information Ecologies

Indigenous information ecology diverges from Western information ecology in a number of significant ways. Initiatives by professional information organizations, including the American Library Association Traditional Cultural Expression policy, and the Society of American Archivists Protocols for Native American Archives are creating policy within this zone of divergence. [7]

Although Western information practice is supported by existing legal and policy formulations, Indigenous customary law and policy has traditionally been defined by non-Indigenous scholars and practitioners. Today there are more than 300 tribal libraries, archives and museums, as well as tribal colleges, policy institutes, and legal institutions that are implementing Indigenous solutions to information challenges. [8] These manifestations of Indigenous praxis show how information impacts tribal sovereignty, language, land claims, traditional knowledge, research protocols, and ethics. [9] Utilizing online, blended learning platforms helps

communities articulate Indigenous information and knowledge management practices.

2.4 Potential Questions

Indigenous methodologies examine the relationships between individuals, communities, places, practices, and technologies. To clarify, we may pursue the following questions:

- How can non-Indigenous researchers utilize Indigenous methodologies?
- How do Indigenous methodologies intersect with issues of diversity, anti-oppression, and other critical theoretical frameworks?
- What are the commonalities between Indigenous methodologies and frameworks offered by community informatics, socio-technical systems analysis, and participatory research methods?

The discussion will introduce participants to ways of thinking about information science research from an Indigenous point of view, as well as ways that information science methodologies might be enhanced by Indigenous methods.

3. REFERENCES

[1] Gershon, N. and Page, W. (2001) What Storytelling Can Do For Information Visualization. Communications of the ACM, 44, 8, 31-37.

[2] Kazmierczak, E. (2001) A Semiotic Perspective on Aesthetic Preferences, Visual Literacy, and Information Design. Information Design Journal, 10, 176-187.

[3] Champagne, D. and Goldberg, C. 2005 Changing the Subject: Individual versus Collective Interests in Indian Country Research. Wicazo sa Review, 20, 1, 49-69.

[4] Smith, L. 1999. Decolonizing Methodologies: Research and Indigenous Peoples. Zed Books, London.

[5] Konkle, M. 2008. Indigenous Ownership and the Emergence of U.S. Liberal Imperialism. American Indian Quarterly, 32, 3, 297-323.

[6] Kamira, R. 2003. Te Mata o te Tai—the Edge of the Tide: Rising Capacity in Information Technology in Aotearoa-New Zealand. Electronic Library, 21, 5, 465-475.

[7] First Archivist Circle. 2006. Protocols for Native American Archival Materials. www.firstarchivistcircle.org/files/index.html Accessed 18 November 2009.

[8] Nakata, M. 2007. Indigenous Digital Collections. Australia Academic and Research Libraries, 38, 2, 99-110.

[9] Holm, T., Pearson, D., and Chavis, B. 2003. Peoplehood: A Model for the Extension of Sovereignty in American Indian Studies. Wicazo sa Review, 18, 1, 7-24.

Roundtable: Sharing experiences related to developing theories for the information field

Martha Garcia-Murillo
Syracuse University
245 Hinds Hall
Syracuse NY 13244
(315) 443-1829
mgarciam@syr.edu

Martin Weiss
School of Information Sciences
720a IS Building
135 North Bellefield Ave, Pittsburgh PA
(412) 624-9430
mbw@pitt.edu

Allen Renear
University of Illinois at Urbana-Champaign
314 LIS, 501 E. Daniel Street MC-493
Champaign, IL 61820
(217) 265-5216
renear@illinois.edu

ABSTRACT
This roundtable will allow academics in information schools the opportunity to share their experiences on theory development, the tools and techniques that they have used to facilitate this theorizing process and the challenges that they face doing so.

Keywords
Theory construction

1. THEORY DEVELOPMENT IN THE INFORMATION FIELD
We, as academics, aim to make contributions to our fields. This is becoming easier as a result of the amount of scholarly work that is now available through digital libraries, databases, and journals. There are, however, some challenges. While much attention is given in graduate education to scientific method and theory testing there are few courses that cover the topic of theory construction/development. As a result many doctoral students and junior faculty do not have a clear idea of how to make a contribution, how to develop their own theories, and how to build and extend the work of others to move a field forward. The process of theory development is more a creative process than is typically taught in methods classes. Theory development improves when we challenge our own assumptions and step out of the comfort of our fields to explore ideas from other disciplines. This is particularly true in the information field which crosses so many disciplines.

2. GUIDING QUESTIONS
To facilitate the discussion in the roundtable we have formulated the following research questions.

- How do you go about finding your contributions?
- What techniques have you used that have made it easier for you to foster your imagination as you try to develop explanations to a problem?
- Have you used any tools that make it easier for you to find, organize, and manage the amount of information now available online?
- How do you find the holes in your research field?
- Have you made any attempts to bring ideas from another discipline that can help formulate an answer to a problem in your field?
- Have you used metaphors as a device to develop theory?
- What methodological tools have made it easier for you to develop your own contributions?
- What are the main obstacles that you have confronted when trying to develop a theory of your own?

This roundtable is the starting point to a conversation that will continue online on a monthly basis. We will decide at the conference the topics that we would like to cover each month.

There are several ways in which these discussions will be facilitated. We will develop a webpage dedicated to theory construction which will identify papers that cover the issue. It will also be the platform for collaborative blog discussions on the topic of theory construction. On a monthly basis we would connect using Adobe Connect to discuss the topics selected during the iConference.

3. ROUNDTABLE LEADERS
The individuals facilitating this roundtable were selected because their work has been on the topic of theory development or because of their interest in supporting research efforts at their academic institutions given their academic responsibilities

Martha García-Murillo
Syracuse University, School of Information Studies
mgarciam@syr.edu

Martin Weiss
University of Pittsburgh/School of Information Sciences
mbw@pitt.edu

Allen Renear
University of Illinois at Urbana-Champaign
renear@illinois.edu

4. REFERENCES

[1] Shoemaker, P. J., Tankard, J. W., & Lasorsa, D. L. (2004). How to build social science theories: Sage Publications

[2] Turner, J. H. (1988). Theory building in sociology: Assessing theoretical cumulation: Sage Publications (CA). Price: 80.71 Length: 152 pgs.

[3] Asher, H. B., & Association, M. P. S. (1984). Theory-building and Data Analysis in the Social Sciences: University of Tennessee Press

[4] Hage, J. (1972). Techniques and problems of theory construction in sociology: John Wiley & Sons.

[5] Davidson-Reynolds, P. (1971). Primer in Theory Construction. Indianapolis, IN: Bobbs-Merrill

Mapping the Intersections of Information Studies and Gender and Sexuality Studies

Patrick Keilty
UCLA
UCLA Department of Information Studies
GSE&IS Building, Box 951520
Los Angeles, CA 90095-1520
011-310-825-8799
pkeilty@gmail.com

Rebecca Dean
UCLA
UCLA Department of Information Studies
GSE&IS Building, Box 951520
Los Angeles, CA 90095-1520
011-310-825-8799
beccadean@ucla.edu

ABSTRACT

The interdisciplinary nexus of information studies, gender, and sexuality studies will be explored in a roundtable format. The discussion will promote ongoing and needed discourse of queer, feminist, gay and lesbian critique in relation to information, digital, and archival paradigm.

General Terms

Documentation, Design, Human Factors, Standardization, Languages, Theory.

Keywords

Gender, Sexuality, Feminism, Queer, Science and Technology, Information, Archives, Libraries, Internet

INTRODUCTION

This roundtable discussion presents research that maps the interdisciplinary space of informational phenomena and institutions alongside the study of gender and sexuality. Such an intellectual convergence can be seen in the growing body of work on affect and archives, as well as information and embodiment. Gender and sexuality research have leveraged a number of information studies concerns, such as surveillance and documentation, into new realms of inquiry and theoretical engagement. Similarly, information scholars attentive to gender and sexuality concerns provide a unique lens on contingent processes of power entangled within our digital and knowledge-based society.

As a roundtable, the thread of this discussion will necessarily follow from the expertise of its participants and the themes of their research. The collective format of a roundtable creates a collaborative discussion around some of the gender and sexuality themes that have emerged lately within information scholarship, including: the body as a repository of knowledge ("bioepistemology"), gender and computer coding, masculinity in "virtual reality," cyberfeminism, technological determinism and gender essentialism, sexual sociability online, sex and the Internet, technical change and gender power relations, and technology as a source and consequence of gender relations.

Developing a Collaborative Sandbox for Digital Library Research

Organizer: Xia Lin, Drexel University

Presenters:

Daqing He, University of Pittsburgh

Lorri Mon, Florida State University

Jeffrey Pomerantz, University of North Carolina

Haozhen Zhao, Drexel University

ABSTRACT

The challenges and issues around digital libraries research are of great interest to the ischool community. Questions such as how to best understand the needs of digital library users, how to identify and develop appropriate information services and resources, how to develop and maintain a community of digital library users, and how to measure the impact of digital libraries to the lives of users, are all central to iShcool's research and education.

As a field of study, digital libraries emphasizes both theories and practice. For many years, iSchools are often focusing more on theories than on practice; an exception is the Internet Public Library (IPL), which was developed originally at the University of Michigan and now is managed and developed at the Drexel University and supported by the IPL Consortium. IPL has been a popular public library with real users from all of the world (the IPL website receives over one million unique visitors per month). It is a digital library designed, implemented and managed by iSchool students and volunteers. It serves as a good example of a digital library environment where students can practice their knowledge and skills.

Looking forward, IPL represents both challenges and opportunities for the iSchool community. Ten years ago, IPL was developed with the advanced ideas and technologies at that time. As the theories and technologies of digital libraries advance, IPL needs to be further developed. Recently, IPL has just went through a major re-design and upgrade, with a new brand name IPL2 (://www.ipl.org). The new design utilizes XML technology and Fedora repository software. It enhances the IPL collections through the merger of LII and IPL data.

To ensure that the development and re-development of IPL is sustainable, we consider it essential to develop IPL as both a practical digital library and a research laboratory. We need to learn from IPL practice to enrich our understanding of digital libraries; we need also to use IPL as a sandbox to develop prototypes and test new ideas to further theories. In other words, IPL should become a collaborative research environment that will help us complete the circle of theories, practice, and research.

We envision the IPL can become

- both a learning and research environment and an operational digital library.
- easy-to-use systems with significant data and digital collections for various digital library activities.
- a platform and a "sandbox" with rich APIs and visual modules where students and faculty can create mashups and test new ideas and new technologies.
- a demonstration space to showcase how research, learning and collaboration take place in a digital space, spanning across time and geographic location.

The goal of this panel is to discuss the challenges and opportunities and explore the ideas and paths that will help us reach this vision. The panel will start with an overview of IPL2 new design and implementation, and then each of the presenters will give a brief description of current research he/she is doing with IPL2. The discussion will then focus on brainstorming how we develop IPL2 to make it a sandbox and a collaborative space for our research and education in digital libraries. Questions to be discussed include: what collections should we develop? What schema and architecture should we adopt and implement? How should we implement the metadata? What APIs are needed for prototype development and testing? How best to collect reference questions and answers for research purposes? How do we collect and evaluate user's needs? How should we develop and grow the user community as well as the researcher community around IPL2? how open and interactive should the digital library be? How much should it allow users to interact with, add to, rate, create, and remix the resources? These and many related questions will open up the dialog between presenters and audiences and we expect the audience to participate actively and contribute their ideas of how to advance IPL2 to become a collaborative research environment for the iSchool community.

Games in the iSchools

Ian MacInnes
Syracuse University
School of Information Studies
Syracuse, NY 13244
1-315-443-4101
imacinne@syr.edu

Andrea H. Tapia
Penn State University
Information Sciences and Technology
University Park, PA 16802-6823
1-814-865-1524
atapia@ist.psu.edu

1. Roundtable Description

Gaming is a rapidly growing new medium with sales surpassing box office and music revenues. Gaming provides a method of interacting with information in ways that static, non-participatory information containers cannot provide.

Libraries, for example, are supporting gaming activities and educators are integrating gaming in new ways. Most young people (and many older ones) are drawn to gaming activities for leisure, sometimes by themselves, but more often, through sharing the same physical or virtual space with others. Gaming, once relegated to the back rooms and basements, is now discussed frequently on the news and at the dinner table. This phenomenon has also captured the attention of iSchool researchers. Some scholars are exploring the information spaces in which gamers live and support their activities. Others explore ways in which gaming can be used to teach traditional skills, while some look at how gaming illustrates new types of information literacy not easily teachable in a traditional lecture-style format. Some support the gaming creation process through working with industry or developing ways that youth programs facilitate game creation and shared experiences. As libraries support gaming activities, researchers are exploring this intersection to understand how gaming can be effectively used.

There is also a growing body of research on gaming relevant to iSchool areas such as information and telecommunications management. Academic conferences including AMCIS and HICSS have had mini-tracks on the subject for many years. There have been several calls for papers for journal special issues on related topics, particularly the growth of online multiplayer environments as a new medium of communication. Virtual worlds, which originated as large scale open ended games, have grown in popularity to the point where they are becoming increasingly mainstream. This trend will strengthen as technological advances make these environments increasingly compelling.

The increasing use of gaming technologies requires greater attention from academia. Examples of topics that can be studied include business models, the digital persona, HCI elements, mobile gaming, online addiction, the purpose and value of recreational gaming in libraries, and virtual item property rights. A number of universities have recognized gaming as an area of high industry and student demand. The iSchools are a natural home for this type of activity but we are currently behind traditional fields such as education, performing arts, engineering, and communication in building research, industry funding, and academic programs.

The goal of this roundtable is to attract iSchool researchers who are exploring gaming research projects. This will be a sharing roundtable, with the hope of allowing researchers to make connections between schools and across disciplines. Those looking to get involved in gaming research are also welcome to attend and discover potential partnerships. Two similar roundtables were held at the 2008 and 2009 iConferences. Both were well attended and identified several collaboration opportunities. The field continues to grow and generate interest in many areas of universities. One goal will be to foster the possibility of creating multi-disciplinary and multi-institutional grant proposals that will allow iSchools to take their turn at the gaming table.

LIS Education and Tribal Libraries, Archives, and Museums Roundtable

Omar Poler
UW-Madison Graduate Student
1334 Williamson St. Apt #1
Madison, WI 53703
1(608)345-9057
ojpoler@yahoo.com

Christina LPW Johnson
UW-Madison Graduate Student
511 West Doty St., Apt. #302
Madison, WI 53703
1(920)559-7237
clpwjohnson@gmail.com

Catherine Phan
UW Digital Collections
431 Memorial Library
Madison, WI 53706
1(608)265-3059
cphan@library.wisc.edu

Abstract
This roundtable discussion will explore the unique value of incorporating Indigenous information issues in LIS education. Moreover, the discussion will elicit participants' own experience in LIS education while raising awareness regarding this important topic. The discussion will provide a forum for participants to network, share ideas for advancing Indigenous information in iSchools, and consider how the study of tribal cultural institutions and knowledge often forces a reassessment of traditional LIS education.

Keywords
LIS Education, Tribal Libraries, Indigenous Information Studies

1. Background & Topic
Since April 2008, University of Wisconsin-Madison School of Library and Information Studies (SLIS) graduate students have increasingly engaged in Indigenous information issues, a highly overlooked area of library and information studies. Since the fall of that year, the motivated students have developed a new and innovative course on tribal cultural institutions. Known as the Tribal Libraries, Archives, and Museums (TLAM) project, the class provides an overview of, and most importantly, practical experience with current issues facing tribal cultural institutions, as well as their history and development.

TLAM began as an independent study project of three SLIS students. Through funding from both the Kauffman Entrepreneurship Community Internship Program and UW-Madison SLIS, the students completed an interest and priorities assessment for the Red Cliff Ojibwe community to determine the potential future role of their library. Through the work in this initial project, the students, along with new recruits, have continued the experience by designing a new course for SLIS. Through this statewide project, the students have broadened their education through experiential learning and have sought to create connections and relationships not only among UW-Madison and the Red Cliff Ojibwe, but also the library and information studies community and several of the eleven federally recognized tribes and bands throughout Wisconsin.

The UW-Madison TLAM course challenges traditional LIS education in a number of ways. It looks beyond the University to tribal librarians, archivists, museum curators, and other community leaders as highly qualified educators on the topic. It is a non-hierarchical network of knowledge seekers teaching one another; there is no "one teacher," but rather it relies on the contributions of all of its participants—students and professionals alike. It is organized around a combination of participatory classroom gatherings, guest speakers, and local events relevant to the course. Experience is especially emphasized through visiting communities, their tribal cultural institutions, and spending time with tribal members. Overall, it is structured as a service learning experience that seeks to create mutually beneficial relationships between tribal communities and LIS. TLAM continues to be offered at UW-Madison SLIS and is an ongoing development.

This roundtable discussion will explore the unique value of incorporating Indigenous information issues in LIS education. Moreover, the discussion will elicit participants' own experience in LIS education while raising awareness regarding this important topic. The discussion will provide a forum for participants to network, share ideas for advancing Indigenous information in iSchools, and consider how the study of tribal cultural institutions and knowledge often forces a reassessment of traditional LIS education. Relationship-building with tribal cultural institutions often also provides an important opportunity for student internships and service learning that can result in mutually beneficial skill sharing.

2. Possible Questions
After a brief presentation on our own participation in Indigenous information issues, there will be a roundtable discussion with attendees, which will likely include the following questions:

• Have you worked with tribal communities?

• In your experience, has your LIS program engaged with tribal communities and/or incorporated Indigenous information issues?

• How can we advocate for tribal cultural institutions within LIS education?

• What resources might help LIS schools develop their own course on tribal archives, libraries, and museums?

• What resources might help LIS schools develop continuing education programming and courses for current tribal cultural institution practitioners?

• What outreach and recruitment efforts might be most effective for building and maintaining networks around this topic?

• Are there additional opportunities for tribal communities and LIS schools to collaborate?

3. Post-Conference Follow-Up

Interested participants will be asked to engage in a network of students, scholars, practitioners, and other interested persons working to inspire and better incorporate the presence of tribal cultural institutions and indigenous information issues in LIS curricula. A few of the opportunities for follow-up will include working with the American Indian Library Association (AILA), becoming AILA members, and encouraging participation in national tribal archives, libraries, and museums conferences. Participants may also opt to be added to our TLAM network database.

Social Networking and Long-Term Organizational Goals

Maeve Reilly
iSchools
501 E. Daniel
Champaign, IL 61820
(217) 244-7316
mjreilly@illinois.edu

Anthony J. Rotolo
School of Information
Syracuse University
245 Hinds Hall
Syracuse, NY
(315) 443-3409
anrotolo@syr.edu

Sherry L.K. Main
Donald Bren School of Information and Computer Sciences
University of California, Irvine
6056 Donald Bren Hall
Irvine, CA 92697-3425
(949) 824-1562
sherry@ics.uci.edu

Richard Urban
Graduate School of Library and Information Science
University of Illinois
501 E. Daniel
Champaign, IL 61820
(217) 443-3409
rjurban@illinois.edu

ABSTRACT

Social and new media tools have increased in popularity and accessibility over the past year and have become much more than social networks and entertainment sites. As students and the press alike come to rely on social media as a real-time information source, how can the iSchools integrate new media tools to create a seamless, integrated flow in both its communications and marketing as well as in its teaching and research tools?

Categories and Subject Descriptors

Information seeking and use, Information management, Social tagging, Social networks, Organizational informatics

General Terms

Performance, Design, Experimentation, Human Factors

Keywords

Social networking, communications, new media

1. INTRODUCTION

In this session we will explore what it means to use social media. We?ll explore topics such as:

* How do I brand my social media presence?

* How do the different social and new media accounts integrate? from Facebook to Twitter and LinkedIn, to YouTube, Flickr and more, we will investigate how to use various tools to complement your message.

* How easily can I develop a social media strategy that aligns with my organizational goals, and at what cost?

* How can my organization integrate social media internally to enhance communication and student services?

* How can I incorporate social media tools in my classroom or in my interactions with students or colleagues across institutions?

This session relies on the expertise of the social media stratetgist at Syracuse, and iSchools communications specialists who will discuss how social networking can become part of the iSchools long-term organizational goals. A graduate student in one of the iSchools will discuss how students perceive these messages, and how researchers and scholars are incorporating social media tools as they procure funding, develop research goals and further the mission of the iSchools.

This roundtable on the practice of incorporating social networking will benefit iSchools communications staff, as well as those about to embark in jobs in the field.

Roundtable:
Research Methods in Community Informatics at the Broadband Moment

This roundtable will present and explore a range of research methods for community informatics. As a multidisciplinary field, community informatics methods are varied and complementary. As an emerging field with fuzzy boundaries, community informatics can benefit from debate and cross fertilization with research that examines similar phenomena, but does not call itself community informatics.

Up to now, the dominant approach in community informatics has been the ethnographic case study method, participant or involved observation. A small number of scholars are using archives and other documents and records, retrieving or repurposing data for studies of multiple cases. Still others use surveys, interviews, and government datasets to examine a large number of cases.

February 2010 is a special moment to reflect on methodology for community informatics. Close to $7.2 billion in stimulus spending is being awarded to bring broadband, public computer centers, and broadband adoption programs to local communities. A National Broadband Plan is to be announced this month as well. This is expected to result in an explosion of community informatics activity which could be the subject of expanded community informatics research. The U.S. Departments of Commerce and Agriculture and the Federal Communications Commission are debating what data to collect from this activity, what data to share and how. They are searching for ways to encourage research to inform the nation's technology policy. A national broadband plan is to be announced in February. Community informatics may be able to make an advance in the wake of these developments if we can pool our methodological knowhow and turn up the power of our microscopes.

This moment has an echo in the origins of community informatics. In the 1990s, community informatics rode a wave of investment in community technology that came from federal agencies, philanthropies, and (primarily hi-tech) corporations. Today the national broadband policy discussion includes debate over what methods and metrics should guide research. Community informatics in the I-Schools has a role to play in this debate.

Six scholars will describe the methods they have used in order to answer the fundamental question "How do communities compute?" The focus here is primarily but not only on local communities and communities of interest within those local communities. There are three aspects to this question:

> What is community? How is it measured?
> What is informatics? What are the metrics of access and use of digital tools?
> What theories explain these data?

There are two goals for the panel:

1 Methodological breadth: We'll consider diverse types of data and explore how standardization could lead to greater comparability.
2 Methodological depth: We'll consider the utility of multiple research methods to examine the same case or sample of cases.

Six i-school faculty-scholars will participate on the panel, and the chair will be Allen Renear, Associate Dean for Research at the University of Illinois.

1 **Lynette Kvasny, Penn State University**. Lynette Kvasny will discuss ethical dilemmas that were raised during her CI fieldwork with a small business association in West Philadelphia.
2 **Chris Coward, University of Washington**. Chris will describe a multiple methods approach currently being employed for a 5-year study to assess the social and economic impacts of public access computing across eight countries. Methods include national inventories and classification of centers, operator and user surveys, and a series of in-depth studies that use ethnographic, quasi-experimental, and other techniques to interrogate specific impact mechanisms.
3 **John Carlo Bertot, University of Maryland**. John Bertot will discuss the survey methodologies that he has practiced in examining public computing in libraries since 1994 as well as the field methods that have complemented and extended that work.
4 **Mia Lustria, Florida State University.** Mia Lustria will talk about a 3-year grant to develop and evaluate a tailored reminder system to encourage breast cancer screening among rural, underprivileged women. Her work spans a multi-method approach including participatory approaches involving clinicians and potential beneficiaries in the design of the reminder system and a randomized controlled trial of the prototype in two rural communities in Florida. Endpoints include system use by clinicians, use of mammogram services as well as behavioral intentions of eligible patients.
5 **Noriko Hara, Indiana University**. Noriko Hara will discuss a comparative study that examines tacit knowledge transition in Japan, Singapore and Taiwan using quantitative surveys, qualitative interviews, social network analysis, and time diary.
6 **Kate Williams, University of Illinois**. Kate will describe and compare three studies: a case study of a Toledo community technology center, a repurposing and use of the federal TOP data, and a current multi-method study of Chicago branch libraries.

This 90 minute session will allow up to 8-10 minutes for each speaker and 30-plus minutes for discussion. Follow-up may include correspondence with federal agencies to explore mechanisms for facilitating data flow and research.

Interdisciplinary Research Agenda on Privacy 2.0

Heng Xu
College of Information Sciences and Technology, Pennsylvania State University, University Park
hxu@ist.psu.edu

Sandra Petronio
Department of Communication Studies, Indiana University-Purdue University Indianapolis
petronio@iupui.edu

Xiaolong (Luke) Zhang
College of Information Sciences and Technology, Pennsylvania State University, University Park
lzhang@ist.psu.edu

Anna C. Squicciarini
College of Information Sciences and Technology, Pennsylvania State University, University Park
asquicciarini@ist.psu.edu

ABSTRACT

Online social networks (OSNs) brought the voluntary disclosure of personal information to the mainstream, rendering the potential intrusion of privacy a critical and acute concern. The main objective of this roundtable is to address the need for a paradigm shift in understanding and addressing users' privacy risks in the Web 2.0 environment with a focus on OSNs. The discussion themes are to: (1) deepen the theoretical understanding of privacy in the context of OSNs, (2) identify privacy intervention strategies for users to prevent privacy threats, and (3) promote privacy awareness.

Categories and Subject Descriptors

J.4 [Social and Behavioral Sciences]: Psychology; K.4.4 [Electronic Commerce]: Security; H.1.2 [User/Machine Systems]: Human factors.

General Terms

Algorithms, Management, Measurement, Security, and Human Factors.

Keywords

Privacy, Online Social Networks (OSNs), Privacy Enhancing Technologies (PETs), Privacy Regulations, and Privacy Awareness

DESCRIPTION

The extensive display of personally information by users of online social networks (OSNs) has made privacy concerns particularly salient. A larger volume of user digital footprints could be potentially accessible to the public [6]. OSNs brought the voluntary disclosure of personal data to the mainstream, thus exposing users' published information with potential abuse by online crooks, stalkers, bullies, and even by their own friends [3, 4]. At the same time, despite the presence of some privacy norms and regulations, there are relatively few well-established institutional rules and contracts governing OSNs, which gives rise to opportunism.

The main objective of this roundtable is to address the need for a paradigm shift in understanding and addressing users' privacy risks in the Web 2.0 environment with a focus on OSNs. This discussion panel aims to establish an interdisciplinary research on Privacy 2.0. The discussion themes are to: (1) deepen the theoretical understanding of privacy in the context of OSNs, (2) identify privacy intervention strategies for users to prevent privacy threats, and (3) promote privacy awareness. In reviewing the extant literature on privacy studies, the following controversial issues become apparent:

- Privacy has been researched for more than 100 years in various fields, e.g., law, economics, management, marketing, psychology, and philosophy. And yet, it is widely recognized that as a concept, privacy "is in disarray [and n]obody can articulate what it means" [8, p.477]. The picture of privacy that emerges is fragmented and usually discipline-specific, with concepts, definitions, and relationships that are inconsistent and neither fully developed nor empirically validated. Facing this challenge and the murky conceptual waters, this roundtable attempts to start an interdisciplinary discussion toward an understanding of information privacy in the context of OSNs.
- This roundtable will also highlight a debate in the privacy practice and research: the relative effectiveness of *technological solutions* versus *regulatory solutions* in ensuring consumer privacy [1, 2]. Skepticism about the effectiveness of privacy-enhancing technologies (PETs) and industry self-regulation in protecting privacy has resulted in privacy advocates and consumers clamoring for strong legislations to curtail rampant abuses of personal information. This roundtable seeks to address to this debate by: 1) discussing how users of OSNs can be protected from privacy threats, and 2) discussing the effects of different privacy intervention strategies on addressing privacy risks.
- Increasingly, many organizations invest considerable resources into the security and privacy awareness and training programs to raise user knowledge about safe computing practices. Unfortunately, despite the fact that considerable resources have been invested to design educational materials to teach users not to fall for security attacks, these materials are often ignored by users [5, 7]. Facing this challenge, this roundtable will discuss how to promote users' collective privacy awareness through facilitating collaborative privacy experiences among users and their peers.

In sum, a multidisciplinary approach will be expected to address above challenges highlighted in the current privacy literature. This roundtable discussion will be build upon ideas from discussion leaders in multiple fields such as information systems (Heng Xu), communication (Sandra Petronio), HCI (Xiaolong Zhang), and computer science (Anna Squicciarini) to explore the conceptual underpinnings of privacy in the context of OSNs, identify privacy intervention strategies, and promote user privacy awareness. During the roundtable, the team will establish the scope of the research, prepare a rough research agenda and plan for future research proposal submissions.

REFERENCES

1. Caudill, M.E., and Murphy, E.P. Consumer Online Privacy: Legal and Ethical Issues. *Journal of Public Policy & Marketing*, 19, 1 (2000), 7-19.
2. Culnan, M.J. Protecting Privacy Online: Is Self-Regulation Working? *Journal of Public Policy & Marketing*, 19, 1 (2000), 20-26.
3. Gross, R., and Acquisti, A. Information revelation and privacy in online social networks. *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, Alexandria, VA, 2005.
4. Kelly, S. Identity 'at risk' on Facebook. *BBC News*, 2008.
5. Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L.F., Hong, J., and Nunge, E. Protecting people from phishing: the design and evaluation of an embedded training email system. *Proceedings of the Conference on Human Factors in Computing Systems*, San Jose, California, 2007.
6. Madden, M., Fox, S., Smith, A., and Vitak, J. Digital Footprints: Online identity management and search in the age of transparency. *PEW Internet & American Life Project*. 2007, http://pewresearch.org/pubs/663/digital-footprints.
7. Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., and Nunge, E. Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. *Proceedings of the 2007 Symposium On Usable Privacy and Security*, Pittsburgh, PA 2007, pp. 88-99.
8. Solove, D.J. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154, 3 (2006), 477-560.

iSchools and the DARPA Network Challenge

John Yen
Penn State
College of Information Sciences and Technology
Information Sciences and Technology Building
University Park, PA 16802-6823
(814) 865-6179
jyen@ist.psu.edu

John Unsworth
University of Illinois
Graduate School of Library and Information Sciences
501 E. Daniel
Champaign, IL 61820
(217) 333-3281
gslisdean@ad.uiuc.edu

Martin Weiss
University of Pittsburgh
School of Information Sciences
720a IS Building
135 North Bellefield Avenue
Pittsburgh, PA 15260
(412) 624-9430
mbw@pitt.edu

Nick Giacobe
Penn State
College of Information Sciences and Technology
0029 Recreation Building
University Park, PA 16802-6823
(814) 863-8555
nxg13@ist.psu.edu

Maeve Reilly
iSchools
501 E. Daniel
Champaign, IL 61820
(217) 244-7316
mjreilly@illinois.edu

David Hall
Penn State
College of Information Sciences and Technology
Information Sciences and Technology Building
University Park, PA 16802-6823
(814) 867-2154
dhall@ist.psu.edu

Jeffrey Stanton
Syracuse University
School of Information Studies
316 Hinds Hall
Syracuse, NY 13203
(315) 443-2879
jmstanto@syr.edu

Wade Shumaker
Penn State
College of Information Sciences and Technology
Information Sciences and Technology Building
University Park, PA 16802-6823
(814) 865-7719
wshumaker@ist.psu.edu

Gary Marchionini
University of North Carolina
School of Information and Library Science
100 Manning
Chapel Hill, NC 27599
(919) 966-3611
march@ils.unc.edu

Tony Maslowski
Penn State
College of Information Sciences and Technology
Information Sciences and Technology Building
University Park, PA 16802-6823
(814) 865-6179
anthonymaz16@gmail.com

ABSTRACT

In celebration of the 40th anniversary of the internet, DARPA launched the Network Challenge to explore issues related to social networking, collaboration, and trust. The iSchools viewed this as an excellent opportunity to achieve multiple goals: (1) to conduct a collaborative research project of interests across multiple iSchools, (2) to enhance the visibility of the iSchools, (3) to collect data for future research regarding social networking and extreme events, whether they occur in the physical space or in the cyber space, (4) to participate the challenge to win. The iSchools DARPA Challenge Team was thus formed.

Keywords

Collaborative research, social networking, trust.

1. INTRODUCTION

This panel will discuss the background of this iSchool project and reflect on how events transpired. After considerable discussion, a strategy centered on the creation of a private network of people associated with iSchools was selected. These people were on- and off-campus students, alumni and faculty and staff of the iSchools. The communications technologies included telephone calls, emails and text messages. Messages were collected and validated by a virtual command center; this command center was also tasked with gathering data from the public internet.

The operation of the virtual command center consisted of three subgroups: the physical search group for processing reports from observers, the cyber search group for using crawlers to monitor the cyberspace; and the visualization group that integrates information into a geo-visual display. While there was a significant amount of information gathered in real-time from crawling the web, the information needed to be verified to distinguish correct DARPA balloon locations from fake ones. This required gathering additional information using crawler or human observers, and correlating multiple types of information (text, images, Google Earth) from multiple sources.

With the goal to reflect, enrich, and share the experience of the iSchools-DARPA Challenge team, the panel will discuss questions including, but not limited to, the following:

-What were the background and the motivation for forming the iSchools DARPA Challenge Team?

-What was the process for forming the team? Can the process be used for future i-School projects?

-How was the message crafted and distributed through the network of iSchools?

-What were the similarities and differences between the approach taken by the iSchools versus those taken by other groups (e.g., MIT's effort)?

-How effective was the physical search and the cyber search efforts, and what were the interactions between them?

-How did the team distinguish correct and fake information?

-What data were collected from the DARPA challenge? How can iSchools leverage these data for future research?

-What were the lessons learned from the DARPA Network Challenge? What were implications of these lessons to broader research questions?

Gone today Here tomorrow: assuring access to government information in the digital age

Shinjoung Yeo
University of Illinois at Urbana Champaign
Grad School of Library & Information Science
Champaign, IL 61820
(415) 902-2511
Yeo1@illinois.edu

James R. Jacobs
Stanford University Library
123D Green Library, Stanford University
Stanford, CA 94305
(650) 725-1030
jrjacobs@stanford.edu

Keywords

Government Information, digital preservation, Federal Depository Library Program (FDLP), Government transparency.

Roundtable Abstract

The issues of access to and preservation of government information are critical to the proper functioning of a democracy. Since 1813, there has been a system in place to insure public access to government information through a partnership between the Government Printing Office (GPO) and the hundreds of libraries in the Federal Depository Library Program (FDLP). For 115 years, the Federal Depository Library Program (FDLP) has assured that citizens had access to information by and about their government.[1]

We are now at a critical juncture. For the last 20 years, more and more government information has been available online, but that information has become more and more ephemeral. Approximately 97% of materials disseminated to libraries have an online equivalent, but less and less is controlled or distributed by the Government Printing Office (GPO).[2] An estimated 44% of Web sites that existed in 1998 disappeared within one year. The average life span of a Web site is less than 75 days.[3] "Fugitive" documents -- those within the scope of the FDLP but not collected or distributed by the GPO and therefore not preserved by FDLP libraries -- are a rapidly expanding problem in the digital world. Worse, just as the US government is harnessing its "information economy" and "information society," it has gradually and systematically expanded its efforts to restrict and privatize government information produced by taxpayers' money -- aided in no small part by powerful private economic and political forces. Since 1980 a significant number of government publications and information has been privatized, repackaged, bought and sold in the market place. From 1981 - 1998, the American Library Association published a series called *Less Access to Less Information by and about the U.S. Government*, a chronology of efforts to restrict and privatize government information.[4]

FDLP libraries have been a bulwark against this gradual shift but they are losing the battle against this digital wave. Some within and outside the library community see the FDLP system as a dated model not appropriate in the digital world, but the concept of a peer-to-peer, redundant and distributed digital FDLP system is still one of the most effective ways to preserve, authenticate and provide access to government information. This system will include the continuation of GPO distribution through the concept of digital deposit of government information (GPO distributing digital files just as it has done with other formats for many years) as a key means of preserving government information, giving widespread access to that information through complimentary collections and digital interfaces, and protecting users' privacy in what they read and access online.

While access to government information is at risk due to weakening FDLP system, privatization and the ephemeral nature of the Web, the Obama administration has been pushing the idea of government transparency as one of his administration's key political agenda points. However, there has been little public discussion of public policy on a long term access to and preservation of government information. In addition, spurred by Obama's use of social networking software for his historical campaign, government agencies have started to use commercial social networking software to distribute government information. There is a minefield of issues involved with the current administration's push for transparency embedded in Web2.0 technologies. Government information is in the public domain by law, but what about government information stored in the cloud and not on .gov servers. Who owns the information? Is it possible to preserve and access that information long into the future? Who is going to protect citizens' privacy to access government information? We often celebrate information technologies without considering the

deeper social and political implications but these questions urgently need to be addressed within libraries, academia and public interest groups. There have been attempts to address these issues within library communities and various interest groups but there has been little coherent effort to reach out to the wider academic communities.

Despite the increased democratic potential of digital government information and information technologies in accessing government information, erosion of the current FDLP system, privatization, use of commercial software, cloud computing, and the lack of specific government policies based on public interests have been endangering citizens' access to government information and future of democracy. As more government information is available online, access for users has changed; this round table will discuss the effect of this on users and how librarians and academics might deal with these changes for future generations of citizens attempting to get information from their government. We will also discuss the critical issues involved and the roles that we envision for the various stakeholders to collect, distribute, and preserve government information in the digital age.

It is crucial to engage in critical dialogue with various government information stakeholders -- librarians, scholars, citizens, journalists, technologists, and information activists. We hope that this round table will bring those players together for a critical discussion of the various issues surrounding government information in the digital age.

Questions that will guide/facilitate the discussion

- What are economic and political forces behind digital distribution of government? What are the consequences?
- What kinds of public policies are needed to secure access and preserve to government information?
- What are technical elements needed to assure long-term access to and preservation of digital government information?
- What are the roles of libraries and academia in providing and preserving government information in the digital age? Does the shift in medium change the roles of libraries?
- What are the requirements and conditions to creating a digital FDLP ecosystem? Are there any other models that should be pursued?

Proposed round table participants:

- Cynthia Etkin (cetkin@gpo.gov), Sr. Program Planning Specialist, Office of the Superintendent of Documents at US Government Printing Office (GPO)
- James R. Jacobs (jrjacobs@stanford.edu), Government Information Librarian, Stanford University, and information activist at Free Government information (http://freegovinfo.info)
- Patrice McDermott (pmcdermott@openthegovernment.org), Director, OpenTheGovernment (http://www.openthegovernment.org/article/subarchive/91)
- ShinJoung Yeo (yeo1@illinois.edu), PhD candidate in library and Information Science and Information in Society Fellow, University of Illinois Urbana Champaign

References

[1] US Government Printing Office. 2009. Federal Depository Library Program Strategic Plan, 2009 - 2014 Draft Discussion Document: 04/17/2009. http://www.fdlp.gov/component/docman/doc_download/37-fdlp-stratigic-plan-2009-2014-draft-3?ItemId=45

[2] Ibid.

[3] Guy, M. 2009. What's the average lifespan of a Web page? JISC-PoWR: Preservation of Web Resources: a JISC-sponsored Project. http://jiscpowr.jiscinvolve.org/2009/08/12/whats-the-average-lifespan-of-a-web-page/

[4] American Library Association Washington Office 1981-1998. Less access to less information by and about the U.S. government. http://freegovinfo.info/library/lessaccess

Wildcards

iConference 2010

FEBRUARY 3-6 • UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

iConference 2010 Submission

Topic:

Impact of Community Technology Centers: Fishbowl Discussion of Methods and Findings

Organizers:

Chris Coward, Mike Crandall and Samantha Becker, Technology & Social Change Group (TASCHA), The Information School, University of Washington

Goals:

This wildcard session continues to build an international community of iSchool researchers who focus on issues of information and communication technologies, public access computing, and international development. A highly interdisciplinary, international group, the iSchool conference provides a unique venue for them to meet and share their common research interests, explore potential avenues for collaboration, and identify ways of further networking and disseminating their work.

Topic:

The Internet and computer technology have changed the way people live around the world. Governments, non-governmental institutions and entrepreneurs have invested significant amounts of human and financial resources in public libraries, telecenters, internet cafés, and other forms of public access. Due to resource constraints, shared access forms the dominant mode of access to these technologies in most developing countries. In the United States, virtually every library provides access (often free) to computers and the Internet. As a result, a large proportion of the world's population has access to a place with computer and Internet resources ranging from basic connectivity to a suite of digital resources, databases, networked and virtual services, training, technical assistance, and trained staff.

Despite these investments, however, we do not have clear evidence of the impacts of public access computing on the people who visit them and communities in which these centers operate. As a model, public access to ICTs has experienced success and failure, leading to both reinforcement of the belief that the model should be expanded and strengthened, as well as to claims that public access ICTs are ultimately ineffective.

This wildcard continues discussion from the 2008 and 2009 iSchool Conferences on the growing research agenda around community technology. In 2008 Mike Crandall and Karen Fisher organized "Let's get wild: Building a national research and service agenda for community technologies and networking," which drew over 60 participants and initiated a Wiki discussion that continued to 2009 wildcard organized by Clara Chu et al., which again drew participants from all ranks and built a strong iSchool community around technology and social change.

This year's community technology wildcard will focus on impact, specifically the research methods that are used to study forms of impact, the challenges inherent in doing so (particularly across different geographical, cultural and political landscapes), baseline indicators that can be shared across studies or applications to create metacategories, and how empirical findings have been (or can be) used to inform policy and decision-making, application design, and service delivery.

Format:

The wildcard will be moderated by Mike Crandall, and will open with introductions by Chris Coward and Samantha Becker from the Technology & Social Change Group (TASCHA) of the University of Washington Information School, who are conducting synergistic investigations of the impacts of access to computers and the Internet at libraries and other venues around the world. Upon introducing their studies (abstracts below), highlighting key areas of challenge (5-6 minutes each), the fishbowl will commence with 5 chairs occupied by four preselected participants (excluding the organizers) seated in a circle in the middle of the room. Each participant will introduce him/herself and be asked to comment on the session's first theme (below). All other attendees will sit around in concentric circles. If someone wants to join the discussion, then s/he takes the empty chair and one of the four participants moves away. This process of someone joining and someone leaving continues until the session's time draws near and the moderator is ready to sum-up the discussion and share next steps. To facilitate the fishbowl, each discussant will be instructed to keep his/her remarks to 2 minutes. Questions that participants may address include (though the purpose of the fishbowl is to promote a socially-constructed dialogue):

- What does community technology mean in different geo contexts? What other terms are used for it?
- How can we define "impact" and what types of impact do we consider?
- What methods can we use to study impact, and how do they differ across economic, political and social domains?
- What are the challenges to studying impact?
- What is our current empirical record of the impact of community technology on individuals, families, organizations, and society, over the short and long term? (what are the indicators?)
- What difference does it make when libraries (as opposed to other venues) provide community technology?
- What are the policy implications of community technology?

In the effort of community building, prior to the conference we will use our Wiki to promote the event as well as other media. At the event we will circulate a sign-in sheet to document/promote future dialog with the global community technology community, and the ensuing discussion will be posted on the UW TASCHA website.

Speaker Abstracts:

The Impact of Free Access to Computers and the Internet at Public Libraries in the U.S. (tascha.washington.edu/usimpact/)

This project focuses exclusively on the impact of free access to computers and the Internet at public libraries in the U.S. Working with libraries, users, and communities, and an advisory committee of library leaders, researchers, and representatives from public policy organizations, the study will (1) develop candidate indicators for the impact of free access to computers, the Internet and related services in public libraries, (2) document the positive and/or negative results from the presence or absence of public access computing resources and services in public libraries on the users of those services, and (3) examine the effects of library characteristics on these results. The results of the study are designed to guide decision-making and communicate to policy makers and funders the impacts of free access to computing and the Internet in public libraries. The team is focusing specifically on outcomes in the domains of (1) civic engagement, (2) eCommerce, (3) education, (4) eGovernment, (5) health, (6) employment, and (7) social inclusion, with regard to how they affect individuals, families, and communities in the short, mid, and long-term. To identify key areas of public access computing impact, we employed a mixed method approach involving both quantitative and qualitative analyses. The study involved two stages. The first, jointly funded by the IMLS and the Bill & Melinda Gates Foundation,

comprised two phases of data collection: (1) a national telephone survey (n=1130) that included persons ages 14 and up; and (2) case studies of 4 public libraries involving 319 users, library staff and trustees, staff at community agencies, politicians, non-library PAC providers, and people-on-the-street. The second stage study, funded by the Bill & Melinda Gates Foundation, involved a national web survey with 45,000 users at 400 libraries. These data were integrated with the telephone survey results to provide an extended sample, as well as to allow analysis of the results by library characteristics obtained from the IMLS PLS data. The results will enable libraries to gain deeper insight into how their populations compare with others across the nation, and put data and tools in their hands to use for advocacy purposes.

Global Impact Study of Public Access to Information and Communication Technologies (globalimpactstudy.org).

Global Impact Study of Public Access to Information & Communication Technologies is a five-year, $7.2-million international research project sponsored by the Global Libraries initiative of the Bill & Melinda Gates Foundation and Canada's International Development Research Centre (IDRC). The project is implemented by IDRC in partnership with the Technology & Social Change Group (TASCHA) at the University of Washington Information School.

Through telecentres, libraries, and other emerging models of public access to information and communication technologies (ICT), this project examines impact in a number of areas, including employment and income, education, civic engagement, democracy and governance, cultural and language preservation, and health.

Using longitudinal and comparative research approaches, the project seeks answers regarding the magnitude of these impacts and how to measure them, as well as the relationship between the costs of providing public access to ICT and its benefits.

Teach, Learn, Engage: Reflections on Community Informatics Curriculum Development

Aaisha Nafeesah Haykal
MLS student, University of Illinois at Urbana-Champaign
501 E. Daniel Street
Champaign, IL 61820
1(716) 491-4634
haykal1@illinois.edu

Suzanne Im
MLS student, University of Illinois at Urbana-Champaign
501 E. Daniel Street
Champaign, IL 61820
1(951) 259-1303
im6@illinois.edu

Aiko Takazawa
PhD Student, University of Illinois at Urbana-Champaign
501 E. Daniel Street
Champaign, IL 61820
1(734) 272-1003
aikot@illinois.edu

ABSTRACT
This paper describes the 2010 iConference wildcard that aims to discuss the expectations, development, and demands of Community Informatics curriculum in light of its status as an emerging area of study.

Categories and Subject Descriptors
H.1.2 [**Information Systems**]: User/Machine Systems – *human factors, human information processing.*

General Terms
Performance, Design, Experimentation, Human Factors, Standardization, Theory.

Keywords
Community Informatics, curriculum, pedagogy, community, students, engagement.

1.INTRODUCTION
Although there is no single Community Informatics (CI) theory, at its foundation Community Informatics originated in response to the advent phenomenon of information and communication technologies and their effects on diverse communities. Michael Gurstein, the Editor in Chief of the *Journal of Community Informatics* and the first to coin the term, describes CI as "the application of ICT to enable and empower community processes."[1] Brian Loader, another early CI scholar who teaches at the University of York, says it is "navigating the interaction between *transformation* as expressed in information technology or IT, and *continuity* as expressed in a local, historical community."[2] Due to the central role of community in the study of CI, curriculum is predicated on the interplay of teaching, learning and engagement. CI encourages engagement to be a circular process, rather than a one-way relationship; thus, both students and communities should gain direct benefits.

As an emerging field of study, it is important to understand how CI is being taught in higher education. Students of CI will be entering the field in in the near future, so hearing about their experiences while they are in school will provide insight about what types of programs and initiatives they will be implementing in their future careers. A few universities have established CI as a formal specialization, while other schools have classes or projects that often have the same goals, but operate under different names. As a result, there is some variation in what students in CI learn at each institution. This raises the question of whether core consistencies should be implemented across institutions with CI curriculum or whether it should be combined with community-specific studies to be a more effective pedagogical approach.

It is important for Community Informatics programs to be in communication with each other so that as pioneers they can spur developments in the field. Furthermore, by exchanging ideas, concerns, and issues that each program is facing in their respective communities and even within their universities, involved institutions can establish a higher level of proficiency in CI education design and practice. Through collaboration, mutual support, and creating networks, the current focus on case studies or different projects can coalesce into a more comprehensive and informative or action-oriented understanding of the body of CI programs.

2.PURPOSE
The panel has three goals: 1) to examine how Community Informatics curriculum is developing in different universities, 2) to examine if the expectations of students and community partners are being met, and 3) to explore possible needs to be addressed within these programs. To obtain first-hand experience on this subject, the panel will feature students participating in CI practice and projects, community partners, and faculty members who teach CI courses.

The organizers, who are students concerned about how the inchoate Community Informatics paradigm fits into future careers that revolve around community engagement and development, such as librarianship and scholarship, created the panel in order to encourage understanding and awareness across CI programs. In addition, this panel will promote pedagogical discourse by sharing the faculties? thought process and challenges as they design degree programs and course syllabi. Furthermore, the exposure to the individual experiences of the community partners will help

[1] In What is community informatics (and why does it matter)? By M. Gurstein, 2007, p. 11.

[2] As quoted in Social networks and social capital: rethinking theory in community informatics by Williams & Durrance, 2008.

paint a picture of potential technology-related needs of a particular community and best practices based on their and the students' perspectives.

3.FORMAT

The panel will begin with moderators presenting a brief history of CI within education, the purpose and format of the panel. The panelists will then briefly introduce themselves (name, institution, position).

A succession of faculty, students and community partners will present on assigned topics and questions, accompanied by PowerPoint slides. Following the presentation, panelists will field questions from the audience regarding the material discussed during the panel or issues that were missed. We are requesting an additional 30 minutes to allow in-depth discussion of this complicated topic. Panelists can raise further questions and dialogue amongst each other and the audience can engage with panelists on the issues. This will facilitate the outcomes we outline at the end of the proposal.

Proposed Schedule	*Alternate Schedule*
Introduction 5 min	Introduction 5 min
Faculty 1 10 min	Faculty 1 10 min
Faculty 2 10 min	Faculty 2 10 min
Student 1 10 min	Student 1 10 min
Student 2 10 min	Student 2 10 min
Community 1 10 min	Community 1 10 min
Community 2 10 min	Community 2 10 min
Discussion 55 min	Discussion 25 min
Total Time: 120 min	Total Time: 90 min

4.QUESTIONS

4.1Faculty

1.How did you develop the curriculum for your courses? What input/theories/models informed your decisions to assign readings, projects, and papers? What theoretical framework and practical skills do you aim to impart to students by the end of each course?

4.2.Students

1.What expectations did you have in regard to coursework, extracurricular activities, research, and professional development that you would be involved with when you started the program? What methods have worked in your experience, and what can be improved upon?

2.What methods did you learn from your coursework that helped to guide you in assessing the needs of the community for the UFL Technology Assistance project, and shape the way it is run?

3.Do you feel the coursework and activities in CI have helped you to learn about the particular needs of diverse communities and how to address them? How are you approaching this matter in the Community Informatics Corps Seminar?

4.3Community Partners

1.How well are the needs of the community being addressed by the respective projects you have partnered with? What can students and faculty do to better serve the community?

2.Do you see the UFL Technology Volunteer Project as being sustainable if the University were to leave the picture? Why or why not?

3.Did you feel you have a voice in the process of developing SisterNet? How do you feel about scholars/students bringing their ideas to the field as opposed to people in the field approaching the university to address specific community needs?

5.PANELISTS

A complete biography of the panelists will be available at the discussion.

5.1Faculty

Kate Williams (Ph.D., University of Michigan School of Information) has been an assistant professor in Community Informatics at GSLIS at the University of Illinois, Urbana-Champaign for two years.

Steven Jackson (Ph.D., Communication and Science Studies, University of California-San Diego) is an Assistant Professor in the School of Information at the University of Michigan and has been involved in assessing the CI program there.

5.2Students

Susan Rodgers is a second-year graduate student at the University of Illinois at Urbana-Champaign and works in the Community Informatics Initiative as a Research Assistant. She is the manager of the approximately 30 Tech Volunteers at the Urbana Free Library (UFL).

Emily Petty Puckett is a second-year master?s student in the School of Information at the University of Michigan, specializing in CI and Library and Information Services. She is the Community Information Corps (CIC) program coordinator.

5.3Community Partners

Debra Lissak is in her third year as Executive Director of the UFL where she has been employed for nearly 30 years. She received her MLIS Science from the University of Illinois, Urbana-Champaign.

Imani Bazzell has worked as a community educator and organizer for over 30 years. She is the founder and director of SisterNet, a local network of African American women committed to the physical, emotional, intellectual and spiritual health of Black women.

5.4Organizers

Aaisha Haykal, Suzanne Im, and Aiko Takazawa will serve as moderator, timekeeper and provide technical/logistical support during the panel.

6.OUTCOMES

Before the conference, panelists will answer a question that will be posted on the Community Informatics Initiative (CII) web space. The conversation that begins there can continue after the panel discussion and provide a means for networking amongst faculty and students. Issues that come up through this forum can be further explored through the establishment of an annual conference where students and faculty of CI and related studies present and share their research and community projects. Additionally, the panel may spur ideas about new courses or assignments. In addition, the research done by Kate Williams and Aiko Takazawa on CI syllabi in various universities can be analyzed and provide another discussion thread. The panel can also be video-recorded and then posted on the GSLIS and/or CII web space. An analytical paper based on the findings in the panel can be written and then submitted to a peer-reviewed journal.

7.ACKNOWLEDGMENTS

We would like to thank Sharon Irish, Chip Bruce, and Jeff Ginger for their advice and support through the proposal process.

8.REFERENCES

[1] Gurstein, M. 2007 What is community informatics (and why does it matter)?. Milan, Italy: Polimetrica.

[2] Williams, K., & Durrance, J. 2008. Social networks and social capital: rethinking theory in community informatics. The Journal of Community Informatics [Online] 4, 3 (Aug 2008). Available: http://ci-journal.net/index.php/ciej/article/view/465/430

No More Lone Rangers: Setting the Research and Education Agenda for Collaborative Information Work in Virtual Environments

M. Kathleen Kern
University of Illinois at Urbana
Graduate School of Library and Information Science
501 E. Daniel Street, MC-493,
Champaign, IL 61820-6211
1 (217) 244-3604
katkern@illinois.edu

Marie L. Radford
Rutgers University
School of Communication and Information
4 Huntington Street, Room 329
New Brunswick, NJ 08901-1071
1 (732) 932-7500 x8233
mradford@rutgers.edu

Joe Sanchez
Rutgers University
School of Communication and Information
Room 205, 192 College Avenue
New Brunswick, NJ 08901-1071
1 (732) 932-7500
sanchezj@rci.rutgers.edu

Lori Mon
Florida State University
College of Information
268 Louis Shores Building
Tallahassee, FL 32306
1 (850) 645-7281
lmon@ci.fsu.edu

Jeffery Pomerantz
University of North Carolina
School of Information and Library Science
CB 3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
1 (919) 962-8064
pomerantz@unc.edu

ABSTRACT

In this Wild Card session, a team of facilitators will lead an open forum to formulate research and teaching agendas around the concern of collaborative information work.

Categories and Subject Descriptors

H.4.3 [**Information Systems Applications**] D.3.3 [**Programming Languages**]: Communications Applications

General Terms

Human Factors,

Keywords

Pedagogy, Communities of practice, Computer-mediated communication, Social networks, Virtual environments, Reference, Collaboration

1. INTRODUCTION

In the last decade, libraries and information centers have formed virtual reference collaborative organizations including statewide, national, and academic consortia. Yet, individual professionals still answer questions primarily as solo operators. Due in part to the economic downturn that reduces available human and other resources, virtual and physical service desks are increasingly staffed by one person. Other models pull the staffing of virtual reference services into offices and away from the reference desk. The old model of on-the-job, over-the-shoulder, apprenticeship learning that was effective in the past is lost with solo staffing or when reference is conducted from the librarian's office or home desktop. These changes in staffing models could increase isolation from peers.

2. REFERENCE AS COLLABORATION

2.1 Importance of Collaboration

Reference as a truly cooperative effort has great potential (see Radford & Vine, in press). It can improve the answers provided to library users by making specific expertise more readily available. Subject specialists are eager to answer more complex questions in their primary areas, so staff satisfaction can also be improved in collaborative situations. Education of new librarians and staff can be enhanced through enabling learning in a real, on-the-job setting that allows for modeling of research strategies and behaviors and point-of-contact feedback.

2.2 Tradition to Innovation

Traditional staffing models, in existence for more than 50 years, stubbornly persist despite the opportunities presented by today's highly networked world and despite evidence that sophisticated library users want more access to subject specialists and are willing to wait if necessary (Radford & Connaway, 2005-2008). How much are librarians using the collaborative tools (including Web 2.0 applications) available to share expertise among subject experts? Are these tools used for real-time collaboration around a question or as static information exchange? What is required beyond these tools to move library practitioners to a collaborative model both within and across organizations?

2.3 Collaboration as a Norm

In 2006, Pomerantz wrote about "collaboration as the norm" in reference work, citing the history of double-staffed reference desks, telephone and in-person referrals, and contemporary collaborations in the form of online discussion lists and "question-swapping consortia." These are all important ways of obtaining answers to a user's question. What is noticeable in the online collaboration examples is that they lack the synchronous involvement of multiple librarians. Is a mere referral to another librarian or information center a true collaboration? McKenzie (2003) reported on users' perceptions of collaborations between library staff, with many users noting positive outcomes. Quinn (2001) explored ways of fostering cooperation. Both Quinn and McKenzie focused on physical reference desks with multiple information professionals. This, however, is a disappearing model. Can the profession move beyond the hand-off model and use online tools and virtual environments that provide a richer learning environment not only for clients, but also for information professionals?

3. GOALS OF THIS FORUM

This interactive forum will discuss opportunities and challenges for formal and informal work collaborations across different virtual information environments, including live chat reference, Instant Messaging and Text Messaging services, and virtual worlds such as Second Life. Together, participants will identify an agenda for research into the collaborative work of information professionals.

In order for collaboration to be an ingrained behavior, enculturation into this model needs to start early in the professional career. How can iSchools better promote collaboration not only in formal team environments, but also for less traditional information and question sharing? If the client's question is our ultimate concern, information professionals must give up a personal sense of ownership of a question at times, but be rewarded with higher user and personal satisfaction when their subject expertise is tapped by others. How do we encourage students to both learn to be skilled researchers and to be willing to ask for assistance?

There is a growing body of research on the effectiveness of virtual teams in the global corporate world, but this model has not yet taken hold in library environments. How large and fluid can a work team be? Can teams or partnerships be encouraged to form dynamically and across organizations? The public and academic library environments are non-competitive, so why does a "Lone Ranger" mentality persist (Radford, 2009)?

These are among the critical questions to be posed by the facilitators, who have conducted extensive research in the areas of virtual reference, hybrid reference models, and virtual collaboration in higher education. Additional questions and the development of directions for research and action will emerge from the participation of the forum attendees.

4. REFERENCES

[1] McKenzie, P. (2003). "User Perspectives on Staff Cooperation during Reference Transaction." The Reference Librarian, 40(83), 5-22.

[2] Pomerantz, J. (2006). "Collaborative Reference Work in the Blogosphere." Reference Services Review, 34(2), 200-212

[3] Pomerantz, J. (2006). "Collaboration as the Norm in Reference Work." Reference and User Services Quarterly, 46(1), 45-55.

[4] Quinn, B. (2001) ''Cooperation and Competition at the Reference Desk." The Reference Librarian . 34(72), 65-82.

[5] Radford, M. L. (Winter, 2009). "A Personal Choice: Reference Service Excellence." Reference and User Services Quarterly, 48(2), 108-115.

[6] Radford, M. L., & Connaway, L. S. (2005-2008). "Seeking Synchronicity: Evaluating Virtual Reference Services from User, Non-User, and Librarian Perspectives." Grant funded by the Institute for Museum and Library Services, Rutgers, the State University of New Jersey, and OCLC, Inc. Available: http://www.oclc.org/research/activities/synchronicity/default.htm <access date: January 8, 2010>

[7] Radford, M. L., & Vine, S. (in press, February 2010). "An Exploration of the Hybrid Service Model: Keeping What Works." In Diane Zabel (ed.). Reference Reborn: Breathing new Life into Public Services Librarianship. Littleton, CO: Libraries Unlimited.

Next Generation Teaching and Learning – Technologies and Trends

Moderator:

Erin Knight
Masters Candidate 2010
School of Information
University of California - Berkeley
eknight@ischool.berkeley.edu

Panel:

Charles Severance
Clinical Assistant Professor
University of Michigan iSchool

Christine Borgman
Professor & Presidential Chair
Department of Information Studies,
UCLA

George Kroner
Developer Relations Engineer
Blackboard Inc.
Penn State iSchool Alumni

ABSTRACT

The landscape of teaching and learning has been radically shifted in the last 15 years by the advent of web technologies, which enabled the emergence of Learning Management Systems (LMS). These systems changed the educational paradigm by extending the classroom borders, capturing and persisting course content and giving instructors more flexibility and access to students and other resources. However, they also constrained and limited the evolution of teaching and learning by imposing a traditional, instructional framework. With the advent of Web 2.0 technologies, participation and collaboration have become predominant experiences on the Web. The teaching and learning community, as a whole, has been late to capitalize on these technologies in the classroom. Part of this trend is due to constraints in the technology (LMS), and part is due to the fact that participatory media tools require an additional shift in educational paradigms, from instructional, on-the-pulpit type of teaching, to a student-centered, adaptive environment where students can contribute to the course material and learn from one another. This panel will discuss the next generation of teaching and learning, involving more lightweight, modular systems to empower instructors to be flexible, explore new student-centered paradigms, and plug and play tools as needed. We will also discuss how the iSchools are and should be increasingly involved in studying these new forms, formulating best practices and supporting the needs of teachers as they move toward more collaborative learning environments.

Categories and Subject Descriptors

K.3.1 [**Computers and Education**] : Computer Uses in Education - *Collaborative learning*

General Terms

Management, Measurement, Design, Human Factors, Standardization, Theory

Keywords

Teaching and learning, education technology, social media, participatory media

INTRODUCTION

The landscape of teaching and learning has been radically shifted in the last 15 years by the advent of web technologies, which enabled the emergence of the Learning Management System (LMS). LMS providers such as Blackboard, WebCT, Sakai and Moodle provide platforms for managing course content and creating anytime, anywhere access to that content through the network. These systems changed the educational paradigm by extending the classroom borders, capturing and persisting course content and giving instructors more flexibility and access to students and other resources. However, they also constrained and limited the evolution of teaching and learning by imposing a traditional, instructional framework. Each LMS simply enabled, and guided instructors to provide ("upload") all the course materials. Students were still seen as the end-users or consumers of the information.

With the advent of Web 2.0 technologies, participation and collaboration have become predominant experiences on the Web. The teaching and learning community, as a whole, has been late to capitalize on these technologies in the classroom, perhaps because of uncertainty around how to incorporate them, or due to constraints imposed by the LMS. But lately, there has been more and more buzz around the potential of these Web 2.0 tools technologies to improve education. More people are exploring how the embedded ideas of user-generated content, network effects of mass participation, openness and low barriers to entry can be applied to traditional education axioms like student engagement, interaction in learning, and student ownership and management of learning. (Mason & Rennie, 2008)

In response to the buzz, most of the LMS providers have begun to incorporate tools such as wikis and blogs, but they are one of many, potentially buried or even disabled tools. Even when the tools are available, the majority of faculty rarely uses them. One study demonstrated that 95% of LMS usage involved a set of five core content management and broadcast communication tools, which fit the instructional paradigm, whereas tools that encourage participation, collaboration and a more student-centered paradigm (Wiki, Discussion Boards/Forums) were not used much. (Hanson & Robson, 2004). Use of participatory tools occurs in the "long-tail" of teaching and learning and is often unsupported and isolated. (Severance, 2009)

Part of this trend stems from the fact that participatory media tools require a shift in educational paradigms, from instructional, on-the-pulpit type of teaching, to a student-centered, adaptive environment where students can contribute to the course material and learn from one another. With this shift, learning is viewed as the building of connections within communities and the active creation of meaning and understanding through participation. Coined "connectivism" by Siemens (2004), this model contrasts the tradition student-as-empty-vessel models. With roots in Papert's (1980) constructionism and Vygotsky's "Zone of proximal development" and apprenticeship models of learning (Rogoff, 1990), social and collaborative learning can enable students to construct a deeper understanding of material and lead to outcomes not possible in a strictly top-down learning environment.

This panel will discuss the next generation of teaching and learning, involving more lightweight, modular systems to empower instructors to be flexible, explore new student-centered paradigms, and plug and play tools as needed. We will also discuss how the iSchools are and should be increasingly involved in studying these new forms, formulating best practices and supporting the needs of teachers as they move toward more collaborative learning environments.

PANEL

Charles Severance (*Clinical Assistant Professor, University of Michigan iSchool and former Executive Director of the Sakai Foundation*)

EXPERTISE: Preparing for the long tail of teaching tools, samples of next generation tools and the standards that will make them a successful option for instructors.

BIO: Charles is currently the IMS GLC Affiliates Coordinator and Clinical Assistant Professor in the School of Information at the University of Michigan. Previously he was the Executive Director of the Sakai Foundation and the Chief Architect of the Sakai Project. Additionally, Charles is the Author of the book High Performance Computing, Second Edition, published by O'Reilly and Associates. He has a background in standards including serving as the vice-chair for the IEEE Posix P1003 standards effort and edited the Standards Column in IEEE Computer Magazine from 1995-1999.

Christine Borgman (*Professor & Presidential Chair, Department of Information Studies, UCLA*)

EXPERTISE: Learning and cyberlearning trends, digital scholarship

BIO: Christine L. Borgman is Professor and Presidential Chair in Information Studies at UCLA. She is the author of more than 180 publications in the fields of information studies, computer science, and communication. Both of her sole-authored monographs, Scholarship in the Digital Age: Information, Infrastructure, and the Internet (MIT Press, 2007) and From Gutenberg to the Global Information Infrastructure: Access to Information in a Networked World (MIT Press, 2000), have won the Best Information Science Book of the Year award from the American Society for Information Science and Technology. She is a lead investigator for the Center for Embedded Networked Systems (CENS), a National Science Foundation Science and Technology Center, where she conducts data practices research. She chaired the Task Force on Cyberlearning for the NSF, whose report, Fostering Learning in the Networked World, was released in July, 2008.

George Kroner (*Developer Relations Engineer, Blackboard; Penn State iSchool undergraduate alumni*)

EXPERTISE: Experience from inside the largest commercial LMS platform provider, what next generation Blackboard will look like, open source education tools being developed by Blackboard's developer community, commercial educational tools that plug into Blackboard products.

BIO: A member of Penn State's 3rd class of iSchool students, George joined Blackboard following graduation to pursue his career interest in educational technology. At first a technical consultant, he now oversees a community of almost 2,000 developers who create add-ons, plugins, integrations, and customizations for Blackboard's learning platform that extend its base functionality. His efforts ensure that third party developers, be they clients themselves or partners, receive the support they need to develop meaningful tools that can scale successfully to thousands of users. An avid educational tool developer himself, he continues to contribute code to open source Blackboard plugins as well as manage a global yearly professional development program for Blackboard developers featuring conferences in the North American, European, and Asia Pacific regions.

Note: We are also exploring the option of a fourth panel representative, so we may have an additional perspective to add to the already rich set of expertise and viewpoints.

MODERATOR

Erin Knight (*Masters Candidate at the School of Information, UC - Berkeley*)

BIO: Erin is a Masters candidate at the UC Berkeley School of Information after 8+ years working in the education technology realm. Most of that time was spent at Blackboard, the largest LMS provider in the United States, providing a first-hand view into the inner workings of educational technology development, adoption and limitations. Her last stint there was on the Blackboard Beyond team, developing modular (free) social media plug-ins for the Blackboard Learning System, an experience which sparked a passion and dedication to open, social learning solutions. As a final thesis project at the School of Information, she and her colleague, Nathan Gandomi, are exploring participatory media for education to better understand effectiveness, learning outcomes and pedagogical practices involved**.**

SUBMISSION AUTHORS

Erin Knight (*Masters Candidate at the School of Information, UC - Berkeley*)

Nathan Gandomi (*Masters Candidate at the School of Information, UC - Berkeley*)

FORMAT

The moderator will guide the discussion by providing a question, topic or trend and the panelists will then share their thoughts and experiences around the topic/trend. The panel will be conducted in this manner through several topics, and then we will open it up to the audience for an in-depth Q&A. The goals are to get different perspectives on each topic, as well as identify areas for the iSchools to explore further and take leadership on.

REFERENCES AND CITATIONS

Hanson, P., & Robson, R. (2004). Evaluating course management technology: A pilot study. *Educause Center for Applied Research, Research Bulletin*, (24), Boulder, CO: EDUCAUSE. http://www.educause.edu/library/ERB0424

Kvavik, R., & Caruso, J. (2005). Study of students and information technology: Convenience, connection, control and learning (Vol. 6). Boulder, CO.: *Educause Center for Applied Research*, Research Study. http://www.educause.edu/apps/er/erm08/erm0740.asp

Mason, R., & Rennie, F. (2008). *E-learning and Social Networking Handbook*. New York: Routledge.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY: Basic Books.

Rogoff, B. (1990). *Apprenticeship in Thinking*. New York, NY: Oxford University Press.

Severance, C. (2009). Preparing for the Long Tail of Teaching and Learning Tools. *Submitted to International Conference of the Learning Sciences 2010*.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International journal of instructional technology and distance learning*, 2(1).

Ethnographies of Large-Scale Systems: How to study distributed, emerging and complex sociotechnical systems

David Ribes and Steven Jackson
Workshop Organizers and Moderators
Communication Culture & Technology - School of Information
Georgetown University - University of Michigan
dr273@georgetown.edu -- sjackso@umich.edu

Marina Jirotka
Oxford eResearch Centre and Computing Laboratory
University of Oxford, UK
marina.jirotka@oerc.ox.ac.uk

Bonnie Nardi
Bren School of Information and Computer Sciences
University of California Irvine
nardi@ics.uci.edu

Susan Leigh Star
School of Information
University of Pittsburgh
slstar@pitt.edu

OBJECTIVES

We have arranged this session to open discussion and sharing around common experiences, approaches and outcomes of ethnographic studies of large-scale systems. The rich and detailed 'thick descriptions' produced by ethnographers have contributed to the understanding of human-computer interactions, meanings and cultures of the digital environment, and practices of technology use. Ethnography has traditionally focused on a site, a geographically localized community or particular workplaces. However, in an increasingly computer mediated and networked world, ethnographers have had to adapt their methods, their sites of investigation and their objects of analysis. We have many shorthands for these difficulties: distribution, scale, heterogeneous expertises, multiple membership, etc. Often these difficulties are precisely what our research attempts to address but only rarely do we give ourselves leeway to discuss how they affect our own practice.

The five participants in this interactive panel will share strategies, problems, and field experiences from their own studies of large scale systems. Exemplars told as experiences, stories and narratives, are ideal devices for capturing and conveying the complexities of real world field research. These exemplars will serve as the material for an open discussion. Our participants were selected both for their *diverse* modes of interface with their objects of study and a *shared* commitment to ethnography and information systems. This includes a range of 'sites' stretching from funding and the policy sphere, to the activities of design and implementation, to studying actual use of production quality systems.

STRUCTURE

The session will begin with participants' brief presentations recounting a single exemplary experience, approach or method in studying large-scale systems and the research questions these activities have generated. The format will be "5:3:1", that is, a five minute presentation using no more than three slides in order to address a single concept, idea or to illustrate a story. We will then open the floor to discussion amongst presenters and with the audience.

The goal for this session is for the experiences and methods themselves to act as common starting points for a collective discussion of ethnographic approaches to large scale systems. Topics will emerge organically from discussion.

This said, below are some of the topics we expect will come to structure the conversation:

- Traditions of ethnography: e.g., historical ethnography, multi-sited ethnography, virtual ethnography and so on
- Multi-method approaches: quantitative and network methods
- Multi-investigator teams
- Comparative studies
- Longitudinal studies
- Venues for communicating approaches and findings back to our colleagues
- Traditions of 'objective' and 'subjective' research
- Methods or best-practices for investigation
- Funding opportunities and dangers
- Developing long-term partnerships

PARTICIPANTS

David Ribes is a faculty member in Georgetown University's Communication, Culture and Technology Program. Trained in sociology and Science Studies at UC San Diego, he completed his post-doc at University of Michigan's School of Information. Throughout his academic career David has been an 'ethnographer of cyberinfrastructure' -- large scale information infrastructure for the sciences. His dissertation research focused on the practical work of participants in the GEON project (cyberinfrastructure for the earth sciences) and since then he has continued his explorations of ethnographic methods for studying large-scale systems.

Steven Jackson is a faculty member in the University of Michigan School of Information and coordinator of the school's Information Policy (IPOL) program. His work addresses policy and practice in large-scale collaborative science, information infrastructure and democratic practice in public sector organizations, and analyses of the design and use of information technologies in international development settings.

Marina Jirotka is Director of the Centre for Requirements Engineering, Associate Director of the Oxford e-Research Centre and Associate Researcher of the Oxford Internet Institute. Her research interests have long been concerned with bringing a richer comprehension of socially organized work practice into the process of engineering technological systems with a focus on supporting everyday work and interaction.

Early on in her career she developed the use of video-based ethnographic research for use in Requirements Engineering. This work was done in collaboration with BT and helped solve problems for City of London trading rooms, service centers and control rooms.

Bonnie Nardi is a faculty member in the Bren School of Information and Computer Sciences at UC Irvine. An anthropologist, she is interested in innovative social uses of the Internet. She has studied instant messaging, blogging, and other forms of computer-mediated communication, as well as face to face communication. She is the author of three books, an edited collection and many articles. Her latest book, My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft, will be published by the University of Michigan Press, June 2010.

Susan Leigh Star is Doreen Boyce Chair in Library and Information Science at the School of Information Sciences at the University of Pittsburgh. For many years she has worked with computer and information scientists, with whom she has studied work, practice, organizations, scientific communities and their decisions, and the social/moral aspects of information infrastructure. Her latest book is a co-edited collection with Martha Lampland, entitled Standards and Their Stories: How Standardization, Quantification and Formalization Shape Everyday Life (Cornell U.. Press, 2009).

Posters

iConference 2010

FEBRUARY 3-6 • UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Redefining the Role of Information Brokers: The Case of Ghana's Agricultural Innovation System and Information Communication Technologies (ICTs)

Benjamin Kwasi Addom
Doctoral Candidate
Syracuse University School of Information Studies

SYNOPSIS

The new information and communication technologies (ICTs) are rapidly evolving and continue to transform the modes and patterns of communication by enabling handling of information, facilitating forms of communication among human actors and electronic systems. This has resulted in increasing potentials of intermediary institutions (information brokers) such as libraries, information centers, the traditional agricultural extension services and other development organizations in keeping with their missions to collect, preserve, link, and make available information to those who need it. This intermediary role is critical considering the importance of knowledge and knowledge management approaches in agricultural innovation systems. According to USAID (2003), understanding the place of ICTs in a developing country agriculture depends on four key concepts: i) that knowledge is an increasingly significant factor of production; ii) that all actors in the agricultural sector are part of an evolving Agricultural Knowledge System (AKS); iii) that ICTs accelerate agricultural development by facilitating knowledge management for AKS members; and iv) that ICTs are essential coordinating mechanisms in global trade.

The study used a multi case study approach in three agricultural districts of Ghana to understand the current modes and patterns of communication that exist between and among local farmers', agricultural researchers, agricultural extension agents and other intermediary organizations (information brokers). The study is driven by one main research question - *what is the current state of communication between and among local farmers, agricultural researchers, agricultural extension officers and other intermediary organizations in Ghana?*

The preliminary findings reveal that; i) there is a high production of local knowledge and innovations by farmers from the three study sites; ii) a number of research institutes and universities located within the study sites are also involve in a wide range of global/scientific research relating to agricultural production, processing and marketing; iii) even though the awareness of the potentials of the local innovations by farmers in scientific research and agricultural production is very high among all the actors, very little is being done to take advantage of these; iv) very little has changed over the years in terms of tools and modes of communication being used between and among the actors despite the increasing potentials of the new ICTs; v) a wide range of intermediary organizations (information brokers) have been identified within the system but there is an absence of any formal collaboration among them for effective delivery of services; vi) there is a very weak (if not total absent) linkage between the local knowledge and global/scientific knowledge sources; and vii) there is a maximum

use of local knowledge by the farmers with relatively very high demand for global/scientific information and innovations for improved agricultural production.

PROBLEM STATEMENT

The importance of knowledge *generation, exchange and use* in any agricultural innovation system cannot be overemphasized. Key institutions responsible for these processes include agricultural researchers, farmers, agricultural extension services, and other intermediary organizations. Unfortunately, agricultural knowledge generation has predominantly been the responsibility of agricultural research institutes with little attention to the role of local farmers in knowledge generation. On the other hand, agricultural extension service has also been solely responsible for the transfer of technologies from researchers to farmers in most developing countries. While science and technology has heavily influenced Ghana's agricultural production system for decades now, there is also a huge body of literature on the value of farmers' local knowledge and innovations in Ghana's agriculture (Amanor, 1994). Several studies have revealed that the emphasis is being shifted onto sharing of knowledge between technical experts and local people (Coldevin, 2003) instead of the conventional approach of 'knowledge transfer'. This is being recognized considering the special characteristics of the new ICTs to take knowledge generated from one location to another (Stiglitz, 1999; Colle and Roman, 2003).

Taking into account the rich professional and scientific work that has been going on in these three areas - i) scientific knowledge in agricultural production; ii) local knowledge and farmer innovations; and iii) ICTs for development, one would have expected a synergy for employing ICTs to tap farmers' local knowledge and innovative activities back into scientific research. Empirically, however, little is known (if any) of how access to ICTs in Ghana has influenced the functions of intermediary organizations that act as the main link between the two sources of knowledge – local and global/scientific.

GOAL

The preliminary results of the study being presented through this poster argue that, for a stable and vibrant agricultural innovation system in Ghana, actors need to reconsider ways by which farmers' local knowledge and innovative activities could be incorporated into scientific research for further innovations – a situation that calls for redefining the role of the intermediary organizations. The main goal of the study therefore was to first understand the current situation, and then explore the role of the new ICTs, and how these technologies could facilitate the functions of agricultural research and extension.

CONCEPTUAL FRAMEWORK

The study draws and builds upon knowledge from the following interrelated concepts and fields;

a) The significant contribution of scientific knowledge and innovations to agriculture (Agrawal, 1995; Andersen, 2007) through transfer of technology model (Rogers, 1962); training and visit model (Tanaka, 2007); and farmer field school model (Simpson and Owen, 2002);

b) The value of farmers' local knowledge and innovations in agriculture in developing nations (Amanor, 1994; Kamangira, 1997; Bellon, 2001);

c) The unique characteristics of the new ICTs as invaluable resources for agricultural research (ISNAR, 2003), and the driver of knowledge and information society (Okpaku, 2003; Dahlman and Aubert, 2002); and

d) The process of identifying, documenting, and incorporating farmers' local innovations into scientific research for repackaging for farmers' use. The theory of absorptive capacity - the capability of any system to acquire, assimilate and exploit external knowledge was proposed by Cohen and Levnthal, (1990) and reconceptualized by Zahra and George (2002). Even though the theory has been successfully explored in firms (Cohen and Levinthal, 1990; Zahra and George, 2002); inter-firm collaborations (Stuart, 1998); and within nations (Liu and White, 1997) to understand the outcome, very little is known of the actual process of absorptive capacity.

Therefore using this theory, the process by which intermediary organizations could facilitate the recognition of the value of varied knowledge sources, their acquisition, repackaging and exchange could be understood.

METHODOLOGY

The study is designed as a qualitative multi-case study using semi-structured interviews and focus group discussions for data collection. Three cases were used based on an existing project, and the types of crops being grown at the three study sites. The sites are also known for i) presence of agricultural research institutes, ii) history of agricultural extension work, and iii) extensive farming activities. Respondents included agricultural researchers from universities and research institutes mandated to work on the crops being produced at the study sites, local farmers, staff of Ministry of Food and Agriculture working at the study sites, and other intermediary institutions mentioned by the above three key stakeholders. The full research process was carried out at the first site and then repeated at the other two sites with minor changes to the instruments depending on the situation on the ground. All the interviews and the focus group discussions were digitally recorded and the audio files transcribed. Secondary documents such as policy documents, institutinal websites, newsletters and pamphelts mentioned during the interviews were also included in the data gathering and analysis. Content analysis of the trascripts is being done using content analysis software – Atlas.ti.

References

1. Agrawal, A. (1995). Indigenous and scientific knowledge: some critical comments, IK Monitor 3(3)

2. Amanor, K. S. (1994). Ecological knowledge and the regional economy: environmental management in the Asesewa district of Ghana. Development and Change 25(1): 4167.

3. Andersen, P. (2007). “Agricultural technology dissemination structures and farmers' knowledge seeking strategies”. The Annual Conference of the Norwegian Association for Development Research (NFU), CMI, Bergen, November 5-7

4. Bellon, M. R. (2001). *Participatory Research Methods for Technology Evaluation*: *A Manual for Scientists Working with Farmers*. Mexico, D.F.: CIMMYT.

5. Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on Learning and innovation. *Administrative Science Quarterly.* 35: 128-152

6. Coldevin, G. (2003). Participatory Communication: a Key to Rural Learning Systems. Sustainable Development Department, FAO SD Dimensions http://www.fao.org/docrep/005/y4774e/y4774e06.htm#TopOfPage

7. Colle, R and Roman, R. (2003). “Challenges in the Telecenter Movement,” in Marshall, S., Taylor, W.,& Yu, X (eds.), *Closing the Digital Divide,* Westport, CT: Praeger.

8. Dahlman, C. J., and Aubert, J-E. (2002). China and Knowledge Economy: Seizing the 21st century. World Bank

9. ISNAR (2002). iNARS: ICTs and National Agricultural Research Systems- e-Development at the Grassroots. Workshop Report: An International Workshop Organized by ISNAR and IICD. The Hague, 16-18 December.

10. Kamangira, J. B. (1997). Assessment of soil fertility status using conventional and participatory methods. MSc thesis, Department of Crop Science, Bunda College of Agriculture, University of Malawi. 107 pp

11. Liu, X. and White, R. S. (1997). The relative contributions of foreign technology and domestic inputs to innovation in Chinese manufacturing industries Technovation, 17(3), 119-125

12. Okpaku, J. (ed), (2003). Information and Communication Technologies for

Africa's Development: An Assessment of Progress and Challenges Ahead, UNICT Task Force, New York, pp. 23-24.

13. Rogers, E. M. (1962). *Diffusion of Innovations*. New York, Free Press. 1st edition

14. Simpson, B. M. and Owens, M. (2002). Farmer field schools and the future of agricultural extension in Africa. *Journal of International Agricultural and Extension Education*, 9 (2), 29-36.

15. Stiglitz, J. (1999). Scan Globally, Reinvent Locally: Knowledge Infrastructure and the Localization of Knowledge, A Keynote Address given at the first Global Development Network Conference, Germany

16. Stuart, T. E. (1998). Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry. *Administrative Science Quarterly*, 43: 668-698.

17. Tanaka, H. (2007). The New Paradigm for the Community Forestry Research and the Implication to the Extension System: Lessons Learned from "Farmer Forest Management School"

18. USAID, (2003). Future Directions in Agriculture and Information and Communication Technologies (ICTs). Background Paper. The Academy for Educational Development And Winrock International, February

19. Zahra, S. A. and George, G. (2002). Absorptive capacity: a review and Re-conceptualization, and extension. *Academy of Management Review* 27 (2), 185-203.

Creating Context for User-Generated Tags: An Exploratory Study

Nicole D. Alemanne
Florida State University
142 Collegiate Loop
Tallahassee, FL 32306-2100
nalemanne@fsu.edu

Besiki Stvilia
Florida State University
142 Collegiate Loop
Tallahassee, FL 32306-2100
+1 850 645 7366
bstvilia@fsu.edu

Corinne Jörgensen
Florida State University
142 Collegiate Loop
Tallahassee, FL 32306-2100
+1 850 644 8116
cjorgensen@fsu.edu

ABSTRACT
This exploratory study investigates methods for enhancing Flickr tags as image metadata through the creation of context. Community generated tags from a sample of images in the Library of Congress's (LOC) Flickr photostream were harvested and compared to metadata from related Wikipedia articles. In addition, a content analysis of comments in the LOC photostream was conducted. This informs an exploration of methods of combining user-generated tags with other resources to create richer, contextual metadata for images. In addition, the LOC and Wikipedia subject terms were compared to subject headings from the Thesaurus for Graphic Materials (TGM) to determine whether socially created metadata can be used to enhance a current knowledge organization tool by suggesting new concepts, terms, and relationships.

Categories and Subject Descriptors
H.5.3 [**Information Systems**]: Group and Organization Interfaces – *collaborative computing.*

General Terms
Measurement, Documentation, Languages.

Keywords
Social Tagging, Metadata, Image Description, Cultural Heritage.

1. INTRODUCTION
Researchers have proposed employing user-generated tags to enhance the metadata and descriptions of cultural heritage resources. Research continues to be conducted to determine if such key words might enhance the description and discovery of resources by adding the users' perspective [1]. Another fruitful area of investigation is the relationship of user-generated vocabulary to currently used metadata schemas and ontologies[2].

However, questions remain about the efficacy of user-generated key words for resource discovery as freely developed tags lack a number of the characteristics of key words developed through the use of controlled vocabularies and thesauri. Tagging has a shallow learning curve [3]. There are few usage rules, and experimentation is easy [4]. However, inconsistent vocabulary usage can create problems for resource discovery [5, 6]. In addition, the creator's definition of a tag, the context for the tag, and disambiguation are not available.

In January, 2008, the Library of Congress (LOC) launched a pilot project in which it made two collections of approximately 3,000 historical photographs available on the photo sharing website Flickr and invited the public to interact with the collections through tagging and description. Many of the images included in the pilot lacked in-depth caption information. Flickr offers its users the ability to append tags, comments, and notes to photos in the collections. By the end of October 2008, more than ten million views had been recorded and 67,176 tags had been added by 2,518 unique Flickr accounts [7]. LOC has continued to add collections to the photostream, and there are over 7,500 images in the collections as of mid-November 2009.

2. SCOPE AND PURPOSE
Because user-generated tags do hold promise for description and discovery, research to determine methods for creating context and disambiguation is an essential component in the broader tag-related research efforts. This project investigates a small sample of images that have user-generated tags, and works with outside resources to attempt to establish methods for creating context and disambiguation.

3. RESEARCH DESIGN

This research was conducted through content analysis, "a method of transforming the symbolic content of a document . . . from a qualitative, unsystematic form into a quantitative, systematic form" [8]. The content is systematized through coding, a process through which the elements are placed in a limited number of categories [8]. In particular, this is a conceptual analysis, in which concepts are quantified, categorized, and [9]. Explicit terms were identified and categorized using manifest coding [8, 9].

The project employs several sources of the data: the LOC photostream in Flickr (http://www.flickr.com/photos/library_of_congress/), Wikipedia (http://en.wikipedia.org/wiki/Main_Page), and the Thesaurus for Graphic Materials (TGM) (http://www.loc.gov/rr/print/tgm1/). A purposive sample of ten LOC images was selected, and the researcher identified Wikipedia entries covering the main subject of these images. Images from seven of the nine collections were included (table 1).

Table 1. Distribution of images by LOC Flickr collection

Collection	# Sample Images
News in the 1910s	3
1930s-40s in Color	2
Abraham Lincoln (1809-1865)	1
Baseball Americana	1
Photochrom Travel Views	1
Women Striving Forward, 1910s-1940s	1
World War I Panoramas	1
FSA/OWI Favorites	0
Illustrated Newspaper Supplements	0

LOC tags were harvested for each image, as were keywords from the Wikipedia entries. The LOC tags and Wikipedia key words were compared to determine the incidence of similarity and difference, to determine the efficacy of combining user-generated tags and Wikipedia-generated key words in creating metadata for images, and to investigate if Wikipedia entries might be used to create context and disambiguation for user-generated key words. In addition, a content analysis of comments in the LOC Flickr photostream was employed to explore the idea of comments as a process of collective disambiguation. Finally, the LOC/Wikipedia subject terms were compared to TGM subject headings to determine whether a current controlled vocabulary might accommodate user-supplied metadata.

4. DATA COLLECTION

Two types of data were collected and coded for the project: user-generated tags were harvested from the ten Flickr LOC images, and subject terms were harvested from the ten connected Wikipedia entries. All tags from the Flickr images were retained and used, with the exception of the "Library of Congress" tag attached to each image (this was considered an administrative tag). Wikipedia terms were harvested from the body and from the information box for each entry; all unique terms were retained. In addition, the comments sections of the LOC Flickr images were downloaded for analysis.

The harvested terms were coded by the researcher in order to investigate the incidence of similarity and difference between LOC Flickr tags and Wikipedia terms and the question of whether Wikipedia might entries be used to create context and disambiguation for user-generated tags. Two sets of codes were used—one for the analysis of similarities and differences in the Flickr and Wikipedia terms, and the other to create categories through which to analyze whether Wikipedia terms might be used to create context and disambiguation for user-generated tags.

4.1 Similarities and Differences

To determine the incidence of similarity and difference between LOC Flickr tags and Wikipedia terms, the researcher developed a coding scheme to differentiate between like and different Flickr and Wikipedia terms. Three categories were used for the similarities/differences coding:

- Flickr tags: Terms that are unique to the Flickr user-generated tags—these terms only appear in the Flickr tags, and not in the Wikipedia entries.
- Wikipedia terms: The complementary category to the first—terms that appear in the Wikipedia entries but were not used by Flickr taggers.
- Similar: Terms that appear in both the Flickr tag lists and the Wikipedia links list. Due to the lack of controlled vocabulary, terms that appeared to be similar were included in this category. For example, the terms 'America', 'United States', and 'USA' were considered to be similar terms.

For this process, each image was coded individually. For each image, the researcher started with the first Flickr tag, and compared it to the list of Wikipedia terms to determine the coding, with all Flickr-only terms coded in the Flickr list and similar terms coded in both lists. After the Flickr tags were fully coded, the Wikipedia links that had not been coded as 'similar' were coded as Wikipedia only terms.

4.2 Term Categories

To investigate whether Wikipedia entries might be used to create context and disambiguation for user-generated tags, the researcher developed a set of codes to categorize the terms. For this process, each image was coded individually. The categories were developed to create facets that represent the possible range of ways that users might describe the subject of images.

The terms were coded into four categories:

- Location: Any term that represents a georeferenceable location or a URL. Georeferenceable locations are those that can be established in terms of map projections or coordinate systems [10].
- Name: Any term that is a proper name but is not georeferenceable.

- Time: Any term that represents an individual point in time or a range of dates or times.
- Description: Any term that does not fit into the above categories.

4.3 Other Tasks

Two other tasks were completed to increase contextual understanding of the results. The LOC Flickr photostream includes a comments section that is used to greater and lesser degrees across images. These comments were analyzed to explore the idea of comments as a process of collective disambiguation. The interaction of the commenters was of specific interest for this task. In addition, LOC/Wikipedia subject terms were compared to TGM subject headings to determine whether a current controlled might accommodate user-supplied metadata.

5. REFERENCES

[1] Trant, J. 2009. Studying social tagging and folksonomy: A review and framework. Journal of Digital Information 10, 1 (2009). http://journals.tdl.org/jodi/issue/view/65.

[2] Stvilia, B. and Jörgensen, C. 2009. User-generated collection-level metadata in an online photo-sharing system. Library & Information Science Research 31, 1 (Jan. 2009), 54-65.

[3] Marchetti, A., Tesconi, M., Ronzano, F., Rosella, M., & Minutoli, S. 2007. SemKey: a semantic collaborative tagging system. In Proceedings of the 16th International Worldwide Web Conference (Banff, Alberta, Canada, May 08-12, 2007). http://www2007.org/workshops/paper_45.pdf.

[4] Beckett, D. 2006. Semantics through the tag. In Proceedings of XTech 2006: Building Web 2.0 (Amsterdam, The Netherlands, May 16-19, 2006). http://xtech06.usefulinc.com/schedule/paper/135.

[5] Golder, S. A., & Huberman, B. A. 2006. Usage patterns of collaborative tagging systems. Journal of Information Science 32, 2 (Apr. 2006), 198-208.

[6] Guy, M. and Tonkin, E. 2006. Folksonomies: tidying up tags? D-Lib Magazine 12, 1 (Jan. 2006). http://www.dlib.org/dlib/january06/guy/01guy.html.

[7] Library of Congress. 2008. For the common good: the Library of Congress Flickr pilot project. Library of Congress. http://www.loc.gov/rr/print/flickr_report_final_summary.pdf .

[8] Monette, D. R., Sullivan, T. J., & DeJong, C. R. 2008. Applied social research: a tool for the human services (7th ed.). Brooks/Cole.

[9] Colorado State University. 2009. Writing guides: content Analysis. Colorado State University. http://writing.colostate.edu/guides/research/content/index.cfm.

[10] Hill, L. 2006. Georeferencing. MIT Press.

Aliases and Ambiguity: A case study of gene aliases, and implications for information curation and AI

Chandler Armstrong
University of Illinois, Urbana-Champaign
Department of Sociology
57 Computing Applications Building
605 E Springfield Ave
Champaign, IL 61820
carmstr3@illinois.edu

ABSTRACT

This research seeks to understand how names and aliases of concepts are used in scientific literature. Natural language processing tools, and data curation in general, depend upon unique concept identifiers for information, and aliases only provide more oppurtunity for ambiguiyt; despite this, aliases seem to persist in literature and daily life. As a case study, gene names are analyzed. This article presents a discussion on patterns of alias usage, and implications this has for bioinformatics librarianship. Observation suggests that research scientists in the bio-medical fields think about information organization from their contextual perspective, and organize information to be most applicable to their daily research tasks. Information scientist might think about information from a more generalized perspective, and prefer categorizations that minimize ambiguity. Aliases used by scientists probably emerge for functional reasons, each providing distinct semantic roles, despite that they create ambiguity from an information curation perspective. In light of this, information science must be careful to consider contexual needs of information users, and word sense disambiguation models will become increasingly important to deal with the increasingly complex grammar for talking about concepts in scientific research.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Gene Name Disambiguation

Keywords

Gene Name Disambiguation, Name Ambiguity, Synonyms

1. INTRODUCTION

This research seeks insight into name ambiguity, to better understand why single concepts possesses multiple names, and asks how to resolve these situations. It is motivated by the degree of ambiguity found in the names and aliases of concepts used by everyone on a daily basis, in both professional and casual capacities. In bioinformatics, this ambiguity has garnered a special interest. Gene names and aliases, in particular, present thorny difficulties and are the focus of specialized areas of model building and research [5].

Gene-name ambiguity arises where gene aliases have multiple senses; it can be envisioned as a special case of word sense disambiguation (WSD). Genes always have an official name, and commonly have an alias or set of aliases; coincidentally, aliases often serve as abbreviations for full gene names. Chen et al [3] investigate the names and aliases for mouse genes, and reports that while gene name ambiguity is low at about fourteen percent, the aliases are far more ambiguous, at eighty-five percent. Schumie et al [4] analyzed both abstracts and articles for gene symbol versus gene alias usage, and found that full gene names were used only thirty percent of the time in abstracts, and eighteen percent of the time in full text; the remainder the time only a gene alias or abbreviation is used. Current attempts at disambiguation vary from 77 to 100 percent accuracy, depending on the details of the model and the species to which the gene belongs (with genes found in humans being the most ambiguous) [1][6]. In summary, aliases are used more often than full names, aliases are very ambiguous, and disambiguation attempts, while promising, are lacking in key areas.

The hypothesis was that, over time, a gene would come to posses fewer aliases; with ideally all the usage converging into a single name. This expectation arises from the intuition that having fewer names for a thing make it easier to find all the information about it. In the limited amount of manual observation made, we do not see the hypothesized patterns. Rather, usage of all aliases increases across time. This suggests that each alias possesses some distinct semantic function.

The observed patterns makes sense, although why may not be immediately apparent. We reason that most genes appear in multiple species, or be involved in multiple bio-medical concepts (such as diseases and drugs). Thus, genes have a lot of range, and are interdisciplinary. From the user's

perspective, having multiple aliases eases the taks of finding information on a given gene as it relates to a specific area of research. Following analysis, it was clear that the original hypothesis was formed based upon a certain set of assumptions, and these assumptions simply do not hold for the bio-medical research scientist using and creating gene aliases.

The findings suggest that information science needs to be aware that information users may have different needs which will compete with the best practices of information organization. Multiple names, even at the price of ambiguity and diffusion of information, possibly provide some benefit to information users.

If our observations represent the state of affairs, automated disambiguation models will become increasingly important, since eliminating ambiguity through tightly controlled concept naming may not be possible. Disambiguation models are increasingly dependent on metadata provided in curated databases. Where data-mining and AI once attempted to *learn* a word sense, contemporary models are dependent on *infering* a word sense from other information available in the database. This means data curation may not need to focus on providing unique concept identifiers for everything, but should provide a diversity of information about items.

2. METHOD

Gene names and aliases were taken from the Entrez Gene database. Each gene in this database possesses a unique identification number: its GeneID. The database also includes the set of known aliases for each GeneID. We search the entire collection of abstracts available from Pubmed for all gene names and aliases. The search was a simple deterministic search with some processing to maximize matching. First, the abstracts were tokenized into single, bi, tri, and four grams. This is because gene names are often multiple tokens, so it is important to attempt to match sequences of tokens up to some length. Most gene names and aliases will be found with up to a four gram search, however not all will (eg. "A. Thaliana Receptor Kinase 1" requires a sequence of five tokens to possibly match). Tokenized abstracts were searched without modifying case, and with case of all tokens uppered. This is because occasionally gene names will be written in upper case. Finally, if a token contained a dash, it was tried as a single token with the dash, and as two tokens without it. This is because authors occasionally add dashes to multipart names.

The concept of a 'gene' is not well defined by this research. The Entrez Gene data has been used naively; its list of genes may not all be, indeed, genes. In reality, many components go into making a gene, and Entrez Gene data may consist of any of these components. Therefore, the data we are investigating may be components in a process that includes whole genes, but also RNA, protiens, and enzymes. Therefore, we do not simply assume that each of our datapoints are genes. However, we do define our data as information on named entities directly relevant to the scientific field of gene research.

The results of the search are text files listing each GeneID, the alias, and all PubmedIDs containing that alias. The set of PubmedIDs for each alias can then be used to quickly look up information such as dates and authors. Using this information, aliases are grouped by the GeneID to which they refer, and then each mapped to the dates of the articles containing them. Then GeneID's family of aliases is plotted into bar graphs for frequency of usage by date. These plots allow quick visual inspection for interesting patterns in usage.

The results of the plotting are, literally, hundreds of thousands of plots. To find interesting ones, many are randomly selected and inspected. We are interested in plots that have two or more aliases with at least a total of thirty instantiations of an alias (and preferably many more). Patterns which might provide evidence supporting the original hypothesis were also specifically sought.

3. RESULTS

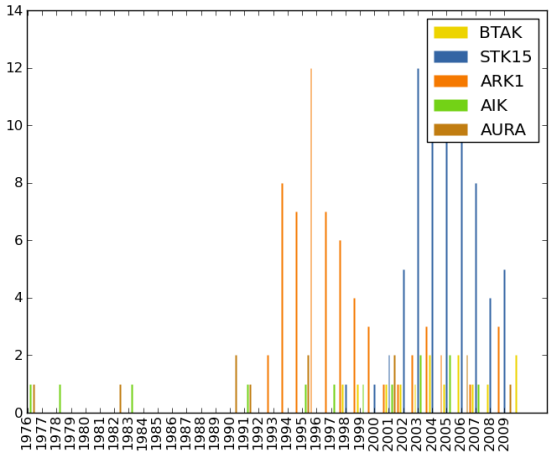
Initially, some plots did seem to support the original hypothesis. However an in-depth analysis revealed that what appeared to be 'advisarial' competition between two aliases was simply an artifact created by ambiguous usage of one of the aliases. One particular case is geneID 6790, the AURKA or Aurora Kinase gene. This gene is present in Humans, and the mold S. Pombe. In S. Pombe the gene is a variant: the Aurora Kinase B. In the literature, when referencing this gene in S. Pombe, the string 'ark1' is most commonly used. 'ark1' is an alias for both Aurora Kinase, and Aurora Kinase B. Additionally this string is a component of 'beta-ark1'; which is not the Aurora Kinase gene (not even the 'Aurora Kinase B' gene, despite the 'beta'). Therefore, a simple match for 'ark1' will often match 'beta-ark1'.

The orange bars in figure 1 below are apparent instantions of GeneID 6790 through the alias 'ark1'. Notice that these instantiations do not align, in terms of variance patterns, with the other aliases; it appears that as the oranga bars decrease, the blue ones are rising. However, the usage seen in the orange bars is almost entirely ambigous: they result from references to 'beta-ark1'. The usage pattern seen in the orange bars is not reflecting how the concept of GeneID 6790 is actually being used. As an aside, when GeneID 6790 (Aurora Kinase) is referenced in Humans, it is usually referred to using 'STK15' (serine/threonine kinase 15), notice that this is the most used alias.

An important element of the ambiguous usage of 'ark1' is that it seemed to always come from other genes. 'ark1' serves as a referent for the gene A. Thaliana Receptor Kinase. More disturbing, 'ark1' is often used in literature to refer to genes who are not listed as possessing that alias. For example, both ARBORKNOX1 and the armadillo repeat-containing protein (an RNA name listed in Entrez Gene) use 'ark1', although this is not listed as a known alias for these concepts. The presence of mispellings and usage of unrecognized names remains a primary source of error in gene name disambiguation models. Current models to resolve this sort of error attempt predict how scientists produce aliases, and generate the likely mispellings and misuses of aliases for a gene [2].

The original hypothesis assumed it is intrinsically beneficial, from an information retrieval perspective, to have fewer names for a given concept. The rejection of the hypothesis

Figure 1: GeneID 6790, frequency of use over time



demonstrates that this is not the case. Rather, it may be the case that it is at least occasionally beneficial to have multiple names for a single concept, in order to represent that concept in a different context. This observation, if true, has important implications for librarianship and data curation. What is ambiguity from the perspective of concept sense, is disambiguity from the perspective of context sense. This also means data-mining and AI models for disambiguating concept senses will become increasingly important.

Many contemporary gene name disambiguation models are based upon data-curation, and using the information provided in the databases (such as author and MeSH terms) to disambiguate gene word sense [1][6]. Because feature selection is swinging away from characterizing the data directly, and instead inferering classifications using other data in the database, errors in the models are likely arising where the integrity of the database breaksdown: with mispellings, unknown aliases, and ambiguity in the curated data (such as author name ambiguity). Improving accuracy of these models must solve these problems, and data curation and ontology research will also become increasingly critical.

Importantly, the problems of ambiguity of concept senses in these databases are interelated. Any name in a database, such as author or journal name, are candidates for ambiguity. Given that gene name disambiguation models may use author or journal of an article as a feature, improving performance of these models depends on disambiguating other named concepts in the database.

4. CONCLUSION

The original hypothesis is a 'natural selection' one in which pressures on the way names are used favors fewer names, and that we might expect to see competitions between names for 'usage resources'. Analysis did not support this hypothesis. Multiple aliases appear to have a function at least some of the time. This has implications for science and research in general, where single concepts are utilized across a wide range of fields, so that it becomes beneficial for the concept to possess a different name for each of these contexts. If this is the case, word sense disambiguation, especially in scientific research, will become increasingly important, as ambiguity will always be rising. The methods and features used by data-mining and AI models are changing to utilize the curation of this mass of data, and utilizing inference from data curation and ontologies to make discoveries and disambiguate word sense.

5. FUTURE

Grouping aliases of a gene and using this as a vehicle for making comparisons and finding relationships has shown to be a productive mechanic. It is likely that other methods of grouping genes may be similarly fruitful. An immediate possibility is grouping the genes of an alias, and using this is a challenge for disambiguating their usage. We found that the nature of gene names appears to produce congruent aliases between many different genes (eg aurora receptor kinase, a. thaliana receptor kinase, beta-adrenegeric receptor kinase, all using the string 'ark1'). Other possibilities include grouping authors of a gene, aliases of genes of a species, journals of a gene, or any other variable which may exhibit a clustering of gene usage, such that certain genes tend to cluster around certain objects (ie around certain authors, journals, disciplines, species, etc).

Given that current source of error in gene sense disambiguation are mispellings and usage of unknown words, models for identifying these anomolies are required. Such models will need to depend even more heavily on data curation and ontologies than ever before, as a word sense may have to be predicted in a set of tokens without utilizing the token itself. Needless to say, this is a difficult problem.

6. REFERENCES

[1] Richard Farkas. The strength of co-authorship in gene name disambiguation. *BMC Bioinformatics*, 9:69, 2008.

[2] Jorg Hakenburg. What's in a gene name? automated refinement of gene name dictionaries. *Biological, translational, and clinical language processing*, pages 153–160, 2007.

[3] C Friedman L Chen, H Lui. Gene name ambiguity of eukaryotic nomenclatures. *Bioinformatics*, 21:2248–2256, 2005.

[4] MJ Schumie, M Weeber, BJ Schijvenaars, EM van Mulligan, CC van der Eijk, R Jelier, B Mons, and JA Kors. Distribution of information in biomedical abstracts and full text publications. *Bioinformatics*, 20:2597–2604, 2004.

[5] Lorraine Tanabe and W John Wilbur. Tagging gene and protein names in biomedical text. *Bioinformatics*, 18:1124–1132, 2002.

[6] Hua Xu, Jung-Wei Fan, George Hripcsak, Eneida A. Mendonca, Marianthi Markatou, and Carol Friedman. Gene symbol disambiguation using knowledge-based profiles. *Bioinformatics*, 23:1015–1022, 2007.

Using History to Study Everyday Information-Seeking Behavior in America: The Case of Car Buying

William Aspray
School of Information
University of Texas at Austin
D8600, Austin TX 78701
1-512-471-3877
bill@ischool.utexas.edu

Barbara Hayes
Indiana University School of Informatics at Indianapolis
535 W. Michigan Suite 475B
1-317-278-7672
bmhayes@iupui.edu

ABSTRACT

Traditional scholarship on everyday information-seeking behavior has focused on approaches that provide a snapshot in time of what is going on in a household. This poster explores the use of history to examine changes over time in both information questions and information sources used in the prosecution of everyday life activities in America. The study is based on identifying endogenous and exogenous forces to the activity at hand, and seeing how these forces cause change. A secondary question raised in this poster is the largely unexamined belief that the Internet has played an exceptional role in changing the nature of everyday information seeking behavior in America. The case of 100 years of car buying in America is used as a particular example, drawn from a larger study of nine everyday American activities.

Categories and Subject Descriptors

K.2 History of Computing, K.4 Computers and Society

General Terms

Economics, Legal Factors, Design, Documentation

Keywords

Information-seeking behavior, information in everyday life, automobile, dealership, safety, environmental movement, suburbanization, war, Depression, advertising

1. The Case of Car Buying

Traditional scholarship on everyday information-seeking behavior has focused on approaches that provide a snapshot in time of what is going on in a household. These include, for example, ethnographic study, reader-reception theory coming out of media studies, and certain kinds of critical theory. (Wellman and Haythornthwaite; Bakardjieva; Lally; de Certeau) Sociologist Andrew Abbott has demonstrated the power that arises from examining a problem of this sort dynamically rather than statically. This research explores the use of history to examine changes over time in both information questions and information sources used when carrying out everyday life activities in America. The study is based on identifying endogenous and exogenous forces to the activity at hand, and seeing how these forces cause change. A secondary question raised in this research is the largely unexamined belief that the Internet has played an exceptional role in changing the nature of everyday information seeking behavior in America.

The case examined here is an examination of car buying over the past 100 years. It is part of a larger study of nine examples of everyday information-seeking behavior: car buying, personal philanthropy, airline travel, genealogy, sports, gourmet cooking, seeking government information, political involvement, comic reading, and text messaging as a youth identity means. (Aspray and Hayes).

The research identifies a class of forces endogenous to the automobile and its industry, which include: advances in technology (such as tires, electric starters, automatic transmissions, and safety features such as safety belts, safety glass, and ABS braking systems); the rise and evolution of the dealership system; the develop of a consumer credit system for purchasing automobiles in the 1920s and for leasing after World War 2; and the rise of foreign competition first from Europe and later from Asia.

The research also identifies a larger class of forces exogenous to automobiles and their industry, which include: the general economic climate in America, especially the Great Depression of the 1930s; the suspension of automobile production during the Second World War; the rise of professional advertising during the same years as the development of the modern automobile as a mass market device instead of being a device for hobbyists; the rise of suburbanization and the growth of women in the workforce, which increased the need for a second car in American households; the environmental movement, which caused people to begin discussions of both pollution created by automobiles and the desire for more fuel-efficient automobiles; the rise of the consumer movement especially since the 1970s, which helped to redress information asymmetries between individuals and corporations; the rise of new media, such as television and the Internet as new venues for gaining information about automobiles; and the development of complementary industries such as oil, road construction, motels, and fast food restaurants.

The purpose of this research is to provide an overview of the information issues and information sources as they changed over time for American car buyers. The number of potential sources is large and the amount of information could easily fill many volumes, so the intention is to be reliable but selective. Before turning to the historical literature, we surveyed the social science literature that informs the car buying decision. For example, there

is an economics literature that discusses the amount of search required to optimize a car purchase decision, and a psychology literature that discusses consumer behavior, often associated with what happens when potential buyers visit a dealership. When examining the secondary historical literature, we sampled from the enormous literature on cars. Sampling this literature helped us to identify general forces and trends that shaped the production and use of cars over the 100 years studied, even though most of this literature is about the supply rather than the demand side.

In order to understand better the demand side, i.e. the way in which owners and potential buyers of cars responded to the changing car environment, we needed to use primary source materials. These included runs of magazines read widely by Americans and archival material about car sales, such as general magazines (Saturday Evening Post), consumer magazines (Consumer Reports), auto enthusiast magazines (Road & Track), and personal finance magazines (Kiplinger's). We examined every car article in long runs of several magazines from each of these types of magazines. Had there been diaries of people's car buying decisions, they would have been useful; but we did not find them.

When choosing magazine articles, it is important to minimize bias in the source material. For example, one might be concerned that if one examines articles from one magazine about the 1930s and a different magazine about the 1960s, some of the differences noted in the articles from these two periods might be an artifact of the different editorial policies and intended readerships of the two magazines. One means for avoiding this research bias is to select from magazines that have long issue runs, such as Saturday Evening Post, which from the 1920s through the 1980s was a major magazine advertising venue of carmakers. All of the primary sources we examined gave only indirect evidence of what questions Americans were asking about cars.

We found in the study that there were some changes in the sources over time. Some sources, such as family and friends, general-purpose magazines, and the local garage mechanic, stayed constant over time. Some other sources became important in the car purchase decision only after the Second World War, such as consumer magazines and hobbyist magazines. Newly introduced media, including radio, television, and the Internet, were employed in advertising automobiles as soon as they became available; and the auto manufacturers were heavy advertisers throughout this entire hundred year period. The Internet was different from the earlier media because it was not entirely controlled by the manufacturers; and indeed the Internet was a powerful leveling source in information asymmetry between individuals and the manufacturers or their dealerships. It is hard to determine what impact the consumer movement played in changing this information asymmetry; it might not only have been the media at play. For example, there are shows on both radio and television today that have the individual car owner in mind more than the interests of the manufacturers in pushing their products.

We found in the study that there were also some changes in the questions asked over time. Some questions were asked throughout this hundred-year period: what is the latest technology? Which cars incorporated it? What was the pricing? Some questions did not arise until later years: questions about fuel economy of cars was not an issue until the 1960s, for until then gasoline was inexpensive and the United States produced more oil than it consumed. Despite smog appearing in Los Angeles in the 1950s, it was not until the 1970s with Rachel Carson's Silent Spring and Ralph Nader's concern about car safety that there was much sensitivity among buyers about environmental or safety issues. A push by Ford in the 1940s to sell cars on their safety features, for example, was a major failure. Local dealer quality did not become a major issue of buyers until the 1960s when the laws went out of effect that limited dealers to selling only in their own geographic region. Quietness was a concern of buyers in the 1920s and the 1990s but not in the era of muscular cars of the 1950s. These are only some examples of the change in questions being raised.

This car study, together with the other case studies mentioned above, provide new information about the nature of exogenous and endogenous forces that shaped information-seeking behavior in everyday American life, the change over time in the questions asked and the sources used, and the reticular role of the Internet as an exogenous force. We will also reflect on the use of history as a tool to advance our understanding of information-seeking behavior in everyday life contexts.

References

Andrew Abbott. The System of Professions (Chicago, 1988).

William Aspray and Barbara M. Hayes, eds. Information in Everyday American Life (MIT, forthcoming).

Maria Bakardjieva, Internet Society: The Internet in Everyday Life (Sage, 2005).

Rachel Carson. Silent Spring (Houghton Mifflin, 1962).

Michel de Certeau, The Practice of Everyday Life (California, 2002).

Elaine Lally, At Home with Computers (Berg, 2002).

Carolyn Haythornthwaite and Barry Wellman, The Internet and Everyday Life (Wiley□Blackwell, 2002)

Annotations and the Digital Humanities Research Cycle: Implications for Personal Information Management

Marie-Eve Bélanger
Faculty of Information, University of Toronto
140 St. George Street
Toronto, Ontario M5S 3G6
+1 (647) 520-0675
me.belanger@utoronto.ca

ABSTRACT
The proposed study assesses the creation, use and organization of annotations in the digital humanities research cycle. It is argued that the gap between digital and physical reading practices creates complex personal information collections, forcing the scholar to cope with information fragmentation by adapting his practices within the constraints of the research process. A poster is proposed outlining a research design and early findings regarding this issue.

Keywords
Annotation, personal information management, scholarly research model

1. INTRODUCTION

Increasingly, primary and secondary textual scholarly resources are available to the humanities scholar in digital form. Such access affords new opportunities for retrieving, exchanging, and storing documents. Recent developments in the digital humanities community have yielded multiple software and web applications supporting the interpretative process of scholars, offering tools for textual and linguistic analysis. However, a large number of academics still find it necessary to print out digital documents to create freeform, idiosyncratic annotations and to profit from the spatiality afforded by the materiality of paper. The available tools and new technologies, as well as the scholar's habits, inevitably create a hybrid information space where information items may be available in multiple formats, organized according to numerous classification schemes, and accessed in various ways. The scholar may then be forced to cope with this information fragmentation by adapting his practices within the constraints of the research process.

In this poster I lay out my reasoning and progress towards understanding the humanities scholar's evolving personal information management (PIM) practices throughout the research cycle. I argue that renewed attention to the diversity of scholarly activity and related annotation practices can inform current research in personal information management.

2. PIM AND THE SCHOLARLY PROCESS IN THE HUMANITIES

PIM refers to the activities performed by an individual in order to "acquire or create, store, organize, maintain, retrieve, use and distribute the information needed to complete tasks" [6]. To accomplish these activities, individuals manipulate information items, physical or digital representations of the information. This study assesses how scholars create, use and organize the information items related to their research project, from ideation to dissemination, explicitly addressing the role of annotations as pervasive information items. This constantly evolving cluster of related information items may be referred to as a personal information collection. While researchers currently disagree on how to correctly define a "collection" [6], this poster aims to provide functional dimensions of personal information collections, hereby adding to the ongoing dialogue in the PIM research community.

The researcher's information collection may change in size and nature throughout the research cycle, according to various phases. While the notion of the humanities scholar as a solitary researcher who values browsing and footnote chaining is now widely understood [3, 14], researchers are still struggling to define the basic, primordial phases of the research process [12, 13, 5, 2, 11, 4]. However, while the terminology, span, and breadth of the research phases vary from model to model, they each touch on common and discrete activities integral to scholarly work across domains, termed "scholarly primitives" by Unsworth [15]. These activities, such as discovering, comparing, referring and sampling, are the basic, constant and recursive modes of interactions between the researcher and his research material.

While these models of scholarly research shed some light on the activities of the scholar and their inherent processes, they fail to account for the variety of annotation practices permeating the various stages of the research cycle. Indeed, most of these models consider the creation and organization of annotation to be strictly relevant to the preparation and elaboration phases of the cycle. However, the typology of reading devised by O'Hara [8] reveals that 1) different types of reading occur throughout the research cycle as seen in Table 01 and, 2) different annotation practices are associated with these reading goals. This is echoed by Palmer and Cragin [11] who hint at the pervasiveness of annotation practices within the scholarship cycle. Building on Unsworth's notion of the scholarly primitive, Palmer and Cragin defined finer grained micro-processes as "information work primitives". Annotation, a type of information work primitive, functions as an articulation device, aligning the different levels, activities and indeed, the different phases of work necessary to the completion of a project. The

Table 1. Comparison of research cycle models and of reading goals using Chu's Research-Phases model as a basis.

Chu (1999)	Stone (1980)	Bradley (2008)	O'Hara (1996)
1. Idea stage	1. Thinking and talking to people		• Reading to learn
2. Preparation Stage	2. Reading what has already been done in the field 3. Studying original sources and making notes	1. Reading and annotating	• Reading for research • Reading to summarize
3. Elaboration Stage		2. Developing interpretation	• Reading for critical review
4. Analysis and writing stage	4. Drafting write-up 5. Revising the final draft		• Reading while writing from multiple sources • Reading to search / reading to answer questions • Proof-reading • Reading for text revision
5. Dissemination Stage		3. Presentation of interpretation	
6. Further writing and dissemination stage			

variety of reading types and consequently of annotation practices may explain why, despite having access to software supporting annotation, academics still tend to print and physically mark up digital documents, rapidly increasing the amount of information items to manage over the course of a project.

3. A FUNCTIONAL TYPOLOGY OF ANNOTATION

As noted above, the personal information collection of the scholar grows both in size and complexity due to the hybridity of the research activities. A functional typology, based on the varying dimensions and formats of annotations may help explain the ongoing gap between digital and paper forms of reading practices and thus characterize the creation, use and organization of information items by the humanities scholar.

Annotations, understood as information items created and manipulated by the researcher, may be used differently according to their content. Several PIM studies, as summarized by Barreau [1] have divided information items according to three categories: archived (long-term value, but unrelated to current work), working (frequently used) and ephemeral (short lifespan and used for non-routine tasks). Boardman and Sasse [1] further refined this classification by proposing four categories: active (working and ephemeral information), dormant (potentially useful but inactive information), not useful and not assessed. This latter categorization may provide valuable insight into the diversity of annotation practices throughout the research process when coupled with the reading goals devised by O'Hara [8] as seen in Table 2. Furthermore, recent research has revealed multiple types of primary uses for annotation: to remember, to think, to clarify, and to share [10]. This poster argues that some of these high-level purposes may be more prevalent than others at various stages of the research cycle.

Types of use as well as organization strategies may also be influenced by the information form of annotations. The choice of tools and medium seems to be related to the formal/informal and explicit/tacit dimensions of annotations, as described by Marshall [7]. Freeform annotations are generally informal and incomplete, demanding to be made quickly and in a minimally disruptive manner. From a cognitive standpoint, paper is the support of choice for readers of printed documents to quickly offload their working memory by creating unself-conscious, informal and incomplete annotations, thus avoiding the loss of information due to an overly disruptive process [9]. In addition, paper may more readily support idiosyncratic annotation methods, such as special signs or individual correction marks, as well as more graphic markings such as margin bars, circling and underlining. While a vast amount of systems developed in the human-computer interaction community may readily support these informal markings by using tablet computers, these technologies have yet to be widely adopted by the digital humanities community, accentuating the fragmentation of information.

4. CONCLUSION

Printing digital documents for reading and annotating purposes leaves scholars with an imbalanced, dual-medium representation. This may increase the complexity of the scholar's personal information collection in 2 distinct ways. The digital and physical copies, unaltered by annotations, represent two separate information items, differentiated by form. These two items are independent from each other, have their own life cycles (one item may be discarded before the other) and are possibly dependent on different organization schemes. Additionally, the content of these items can be independently modified: the printed document, once replete with notes, does not correspond structurally or semantically to its digital version. This poster is centered on the issues emerging from this gap and their effects on personal information management practices of humanities scholars. Early findings, based on a review of the relevant literature as well as interviews and observation sessions of humanities scholars, are reported in the poster.

Table 2. Types of information items created according to reading goals and research cycle phases.

		Phases of the Research Cycle (Chu)				
		Idea	**Preparation**	**Elaboration**	**Analysis/Writing**	**Dissemination**
Annotations made while...	**Reading to learn**	ephemeral	dormant	dormant	dormant	dormant
	Reading for research	-	working	working	working	dormant
	Reading to summarize	-	working	working	working	dormant
	Reading for critical review	-	-	working	working	dormant
	Reading while writing from multiple sources	-	-	-	working	dormant
	Reading to search/reading to answer questions	-	-	-	ephemeral	-
	Proof-reading	-	-	-	ephemeral	-
	Reading for text revision	-	-	-	ephemeral	-

5. ACKNOWLEDGMENTS

The author expresses her gratitude to Dr. Matt Ratto and Dr. Alan Galey for their continued support and constructive comments throughout this research.

6. REFERENCES

[1] Barreau, D. 2008. The persistence of behavior and form in the organization of personal space. *Journal of the American Society for Information Science and Technology*, 59, 2, 307-317

[2] Bradley, J. 2008. Thinking about interpretation: Pliny and scholarship in the humanities. *Literary and Linguistic Computing*, 23, 3, 263-279.

[3] Brockman, W. S., Neumann, L., Palmer, C. L., & Tidline, T. J. 2001. *Scholarly Work in the Humanities and the Evolving Information Environment*. Washington, DC: Digital Library Federation and the Council on Library and Information Resources.

[4] Case, D. O. 1986. Collection and organization of written information by social scientists and humanists: a review and exploratory study. *Journal of Information Science*, 12, 3, 97-104.

[5] Chu, C. M. 1999. Literary critics at work and their information needs: a research-phases model. *Library & Information Science Research*, 21, 2, 247-273.

[6] Jones, W. 2007. Personal Information Management. *Annual Review of Information Science and Technology*, 41, 1, 453-504.

[7] Marshall, C. C. 1998. Toward an ecology of hypertext annotation. in *Proceedings of ACM Hypertext '98*, Pittsburgh, PA (June 20-24, 1998), 40-49.

[8] O'Hara, K. 1996. *Towards a Typology of Reading Goals. Technical Report EPC-1996-107*. Cambridge: Rank Xerox Research Center.

[9] O'Hara, K. P., Taylor, A., Newman, W., & Sellen, A. J. 2002. Understanding the materiality of writing from multiple sources. *International Journal of Human-Computer Studies*, 56, 3, 269-305.

[10] Ovsiannikov, I. A., Arbib, M. A. & Mcneill, T. H. 1999. Annotation technology. *International Journal of Human-Computer Studies*, 50, 4, 329-362.

[11] Palmer, C. L. & Cragin, M. H. 2008. Scholarship and disciplinary practices. *Annual Review of Information Science and Technology*, 42, 1, chapter 5.

[12] Stone, S. 1980. CRUS humanities research programme. In Humanities information research: *Proceedings of a seminar; Sheffield 1980BLR&DD Report No. 5588*, Centre for Research on User Studies, University of Sheffield, Sheffield, England, 15–26.

[13] Stone, S. 1982. Humanities scholars: information needs and uses. *Journal of Documentation*, 38, 4, 292-313.

[14] Toms, E., & O'Brien, H.L. 2008. Understanding the information and communication technology needs of the e-humanist. *Journal of Documentation*, 64, 1, 102-130.

[15] Unsworth, J. 2000. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this? In *Symposium on Humanities Computing: Formal Methods, Experimental Practice*. King's College, London. Retrieved November 1, 2009, from http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html

The force of standards and guidelines in Web accessibility work

Alison Benjamin
Faculty of Information, University of Toronto
140 St. George Street
Toronto, Ontario M5S 3G6
alison.benjamin@utoronto.ca

ABSTRACT
A variety of approaches are taken to address pervasive and persistently occurring barriers to accessibility and inclusion online. W3C Web accessibility guidelines and standards are their infrastructural nexus. This raises the question of what force guidelines and standards have in the work practices of Web developers and designers, who may use these standards and guidelines to a variety of ends.

General Terms
Design, Standardization.

Keywords
Web accessibility, Web adaptability, Inclusive Design.

1.INTRODUCTION
The Web has evolved from an infrastructure oriented around the retrieval of static documents to a platform that supports and mediates social engagement, collaboration, and user contribution of content [10]. Alongside this evolution, a host of barriers to inclusion online have either remained or emerged; so have a variety of strategies to mitigate or eliminate these barriers, encompassing legal, technical, and design-oriented approaches. W3C standards and guidelines are the surface upon which divergent approaches to inclusion connect. There is a difference between a formal standard and its implementation [12], highlighting the analytical importance of context of use and interpretation. A study in-process of Web developers' practices is described.

2.LITERATURE REVIEW
Web accessibility generally refers to the technical approaches used during Web design to make a Website more accessible to users (e.g. disabled people, the elderly) and user agents (e.g. Web browsers, adaptive technologies, and mobile phones) [17]. The World Wide Web Consortium's Web Accessibility Initiative (WAI) provides the Internet's most prominent accessibility regime and prescribes various guidelines and standards that Web content, authoring tool, and user agent developers can use [17]. As a technical standards organization, its guidelines and standards retain a technical orientation [9,5].

According to the WAI notion of accessibility, most Websites are inaccessible [17]. Much research (e.g. [7, 18]) documents WAI implementation gaps.

At the same time, a solely technical approach to Web accessibility – i.e. one based on strict conformance with WAI guidelines and standards – has been criticized by numerous scholars and practitioners. On the theoretical side, the WAI has been critiqued for its foundation in a "universal" design, rule-based, "one size fits all" model of accessibility and new models of Web accessibility and adaptability have been proposed [10,15]. For example, in Inclusive Design approaches, a core problematic is the task of designing for diversity, personalization, and context [10,15].

In practice, a number of online initiatives have emerged that tackle barriers to inclusion online by harnessing social computing strategies, user generated content, distributed peer production, and social networking. These initiatives may constitute a challenge to key aspects of the WAI model. The assumption that Website owners should have autonomy over content and markup is troubled in a social networking/collaborative context where users generate their own markup and content [1,10]. The Fluid Project's (see http://fluidproject.org) user interface components relinquish control of presentation in favor of applying the preferences of users [17]. Its architecture implements ISO 24751, a metadata standard for expressing individual preferences and accessing personalized resources [17]. More generally, the orientation of design and development towards facilitating *participation* in online culture contrasts with the Web Accessibility Initiative model based on accessing an interface "to" a resource [8].

3.THE SIGNIFICANCE OF STANDARDS
Compliance - to one degree or another - with WAI guidelines and standards is a component of all of the above-mentioned approaches to Web accessibility [10]. Bowker and Star's [2] observation that information systems inherit the base inertia of the infrastructures they are installed upon raises the question of what torque these guidelines and standards have in the work practice of designers and developers.

My research is premised on two inter-related assumptions. Firstly, access problematics (e.g. accessibility, adaptability, inclusive design) that have emerged throughout the Web's evolution should not be seen in strictly technological terms because technology itself embodies social relations [6]. It is informed by scholarship that resists "digital divide" debates framed in merely technical terms [4] and individualizing, medicalized accounts of disability [12,15,16].

Secondly, WAI standards and guidelines, as central IT infrastructures that predicate a wide variety of Web design approaches, are socio-technical phenomena [2,13] and must be examined as such. WAI standards and guidelines are central artifacts shaping everyday work practice concerned with access to the Web. In their ongoing repetition they link together an ecology of heterogeneous interests and concerns, as key reference points for industry vendors, law and policy makers, disability rights advocates, and wider discourses about the role and nature of the Web itself.

4.RESEARCH GOALS

To investigate the force of WAI standards and guidelines, my research engages practitioners involved with a range of approaches to Web accessibility. The following research questions are explored:

[1] How is a concept of accessibility explained to designers and developers in the Web Accessibility Initiative? -- texts in question are its key standards: WCAG (Web Content Accessibility Guidelines) & ARIA (Accessible Rich Internet Applications); unit of analysis is the documents themselves.

[2] How are its standards implicated in the work practices/work flows of Web developers and designers?

[3] How are these standards bound up (if at all) with their conceptualizations of access?

Using qualitative data collected from semi-structured interviews, the work practices and standards documents of Web designers and developers will be used as a basis for developing a grounded theory/situational analysis [3] of how accessibility guidelines and standards are implicated overall in development approaches. Key issues are: In what ways are standards and guidelines are seen as relevant (e.g. in terms of making accessibility work noticeable and comparable) and what work goes into facilitating and maintaining them? In the minds of developers and designers, what are the politics of these artifacts: In other words, how are broader social questions about access bound up (if at all) with the use of W3C accessibility infrastructures? Where do users fit in? And what are the non-quantifiable aspects of work in the nebulous task of designing for diverse audiences?

5.CONCLUSION

This study examines how the WAI is implicated in the practices of a variety of practitioners. Interviews are occurring between December 2009 and February 2010. My iConference 2010 poster elaborates on my research design and early results of my study.

6.ACKNOWLEDGMENTS

Thanks to my advisors, participants, and the reviewers of this proposal.

7.REFERENCES

[1] Bigham, J.P., Brudvik, J.T., Leung, J.O. and Ladner, R.E. 2009. Enabling web users to and developers to script accessibility with Accessmonkey. *Disability and Rehabilitation: Assistive Technology 4(4)*, 288-298.

[2] Bowker, G., and Star, S. L. 2000. *Sorting things out: Classification and its consequences*. MIT Press, Cambridge, MA.

[3] Clarke, A. 2005. *Situational Analysis: Grounded Theory after the Postmodern Turn*. Thousand Oaks, California: Sage.

[4] Clement, A. and Shade, L.R. 1998. *The Access Rainbow: Conceptualizing Universal Access to the Information/Communications Infrastructure.* Retrieved on November 10, 2009 from http://archive.iprp.ischool.utoronto.ca/publications/wp/wp10.html

[5] Di Blas, N., Paolini, P. and Speroni, M. 2004. Usable Accessibility" to the Web for Blind Users. In C. Stary and C. Stephanidis (Eds.), *User-centered Interaction* Paradigms *for Universal Access in the Information Society*. Lecture Notes in Computer Science, No 3196. Berlin: Springer.

[6] Frohmann, B. 1993. Communication Technologies and Human Subjectivity: The Politics of Postmodern Information Science. Paper presented at the *Proceedings of the 21st Annual Conference Canadian Association for Information Science* Anigonish, Nova Scotia, 1-14.

[7] Hackett, S. and Parmanto., B. 2005. A longitudinal evaluation of accessibility: Higher education web sites. *Internet Research 15(3)*, 281-294.

[8] Hockema, S.A. and Coppin, P.W. 2009. Beyond the Information Interface. Submitted for review to CHI 2010.

[9] Kelly, B. and Nevile, L. 2008. Web accessibility 3.0: Learning from the past, planning for the future. Retrieved August 1, 2009 from http://www.ukoln.ac.uk/web-focus/papers/addw08/paper-2/html/

[10] Kelly, B., Nevile, L., Sloan, D., Fanou, S., Ellison, R., and Herrod, L. 2009. From web accessibility to web adaptability. *Disability and Rehabilitation: Assistive Technology, 4(4),* 212-226.

[11] Pargman, D. and Palme, J. 2009. ASCII Imperialism. In *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life,* Lampland, M. and Star, S.L. eds. Ithica, London: Cornell Univesrity Press. 177-200.

[12] Shakespeare, T. 2006. The Social Model of Disability. In *The Disability Studies Reader* L. Davis Ed. New York: Routledge. 197-204.

[13] Star, S.L. and Lampland, M. Reckoning with Standards. In *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life,* Lampland, M. and Star, S.L. eds. Ithica, London: Cornell Univesrity Press. 3-24.

[14] Takagi, H., Kawanka, S., Kobayashi, M., Itoh, T., and Asakawa, C. 2008. Social accessibility: Achieving accessibility through collaborative metadata authoring. Paper presented at the *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility,* Halifax, Nova Scotia. 193-200.

[15] Treviranus, J. and Roberts, V. 2006. Inclusive E-Learning. In *International handbook of virtual learning environments*, J. Weiss et al. Eds., Dordrecht: Springer. 439-465.

[16] Treviranus, J. 2009. You say tomato, I say tomato, Let's Not Call the Whole Thing Off: The Challenge of User Experience Design in Distributed Learning Environments. *On the Horizon 17(3)*. 208-217.

[17] World Wide Web Consortium. 2009. Web Accessibility. Retrieved November 16, 2009 from http://www.w3.org/standards/webdesign/accessibility#wai

[18] Zaparyniuk, N., and Montgomerie, C. 2005. The Status of web accessibility of Canadian universities and colleges: A Charter of rights and freedoms issue. *International Journal on E-Learning, 4(2)*. 253-268.

Exploring Impacts on Older Adults' Channel Selection When Faced with an Information Need

Johanna L.H. Birkland
School of Information Studies
Syracuse University
Syracuse, NY 13219
(315) 443-2911
jlbirkla@syr.edu

ABSTRACT

1.1 Introduction: Information Needs and Older Adults

Our everyday lives have become more dependent upon digital information and services. This movement of information to digital formats, which often requires an individual to use advanced technology, is seen as particularly challenging for the older members of our society [20]. Several studies have suggested that the movement to digital information sources puts older adults (those age 65 and greater) at risk of missing the information or the services that they need [20], primarily because this population has the lowest rates of information and communication technology usage [11]. For instance, Medicare is a federally funded program that provides healthcare to older adults in the U.S. However, information about Medicare is only provided by the U.S. government online. If an older adult needs information about Medicare benefits or enrollment, they need to either access official information online, or go to a secondary source [4, 16, 20]. Another salient example of this issue is embodied in the move to electronic voting in the U.S. In areas that have implemented electronic voting, empirical evidence has shown a statistically significant decrease in the amount of older adult voters [12].

Understanding how and why older adults choose a digital or a non-digital information channel when faced with an information need is becoming increasingly important because many developed nations are aging. In the U.S., there is a projected 147% increase in the number of individuals over age 65 from the year 2000 to the year 2050 [15]. With older adults representing a greater portion of the population, it will be important for aging societies to cater to the information needs of this diverse group.

In the literature studying older adults and technology adoption (as well as e-service usage) several factors have been proposed to affect older adult usage of digital information and services. These factors have included relevance/usefulness of the technology to the older adult [14], technological literacy [7, 8], and the perceived security of the digital or non-digital system [10]. Technological literacy, or the basic skills needed to operate a technology, has often been suggested as the largest barrier to older adult usage of advanced technology [7, 13]. However, many older adults have become "post-adopters" of information technologies: they once used these technologies but have stopped [8]. This suggests that other factors are influencing older adults beyond simply knowing how to use a computer. There is most likely a diverse set of factors that influence older adult's likelihood to use a digital information source.

1.2 Contributions of this Study: Exploring Older Adult Channel Selection

This research makes a contribution to understanding why and how an older adult chooses an information channel when faced with a conscious information need. This work applies relevant theory from gerontology, information systems, information science, and communication studies in order to understand the multitude of factors that may influence an older adult's selection of an information channel. By taking a life course perspective, this research attempts to understand how older adults' diverse life experiences impact their likelihood to use a more digital form of an information system (such life impacts could include employment, exposure to technology, and retirement). The Unified Theory of Technology Acceptance [18] is also applied in this research, as several studies have demonstrated that an older adult's perceptions of ease of use and usability have a substantial impact on their technology usage [6, 9, 17, 21]. This research also draws upon gratifications theory, which suggests that individuals become reliant and tend to prefer channels that fulfill their social and psychological needs , rather than choosing channels purely upon their information needs [1, 5].

These relevant theories have been incorporated into a model of channel selection, seen in Figure 1. This model proposes that an older adult goes through a somewhat linear process of information channel selection. This model proposes that an older adult faces an information need- the conscious recognized need for information to complete a task (or complete a step in a task) and/or to answer a question that the older adult has. In order to fulfill this information need, an individual must consult an information channel. Information channels refer to communication and information mediums through which older adults can access information, including mediums such as cell phones, the internet, written pamphlets, etc.

This figure presents a process in which the older adult recognizes their information need (1), chooses a channel (2), gathers information from this channel (3), and then evaluates that

information and the channel in general (4). These evaluations affect a person's likelihood to choose that channel again for a similar information need. This study is particularly interested in the channel selection process, so this part of the model is expanded. The basic design of this model is based upon Wilson's [19] information behavior model and does not assume that an individual is successful in obtaining the information to meet their information need. It may take an individual several attempts at this process in order to be successful, or an individual may choose to end the process at any time.

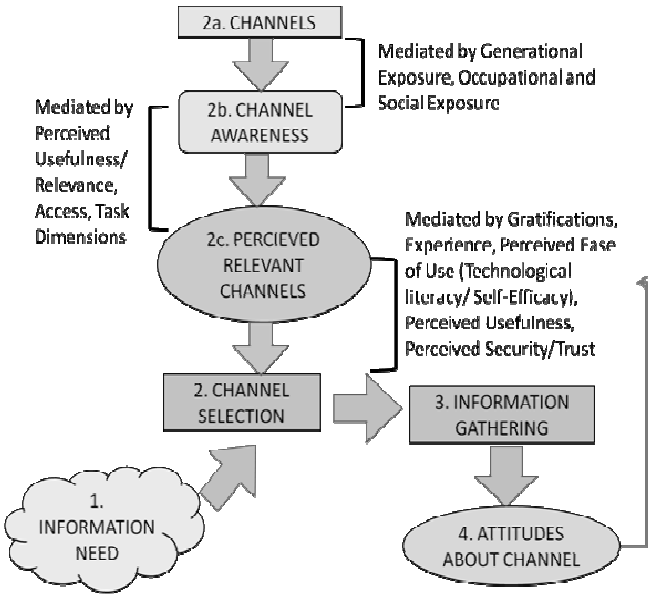


Figure 1. Model of Channel Selection for Older Adults when Faced with an Information Need

The model proposes that certain channels exist in reality (2a)- outside of any particular older adult's perceptions. An older adult, however, is only aware of certain channels. For example a person may not be aware that information can be obtained from YouTube (2b). Channel awareness is mediated by generational, occupational, and social exposure to possible information channels. When faced with an information need, an older adult considers the potential relevant channels that information could be obtained from (2c) based upon their channel awareness. This analysis of potential channels by the older adult is mediated by the channel's perceived usefulness for meeting their information need, older adult's access to that channel, and the task dimensions (such as time constraints, as well as the depth and accuracy of the information needed) involved in meeting that information need. For instance, an older adult may feel that because the internet has relevant information and because the information is needed within a few hours, that the internet may be the best choice. However, the older adult may not have access to the internet, and therefore the older adult decides that the internet is not a relevant choice. An older adult then makes an information channel selection from the potential relevant channels (2). The selection is mediated by the gratifications that an older adult receives from a certain channel, the older adult's perceptions of how easy to use a channel is, how useful that channel will be, and if they believe that the channel is trustworthy and secure.

1.3 Potential Study Design: Exploring Older Adults Information Needs

This study will use a mixed methods multi-phased approach in order to study the factors that influence older adults' choice of media and information channels. The working model proposed above will be expanded and revised based upon the results of the study in order to effectively model the information channel selection of older adults. The study will involve three phases, with each phase building upon the next phase of the study.

During the first phase of the study, older adults will be interviewed in order to understand the types of information needs that are regularly experienced by older adults. From these interviews, a specific set of information needs will be selected (for example, information needs about health issues). The information needs identified will then be used throughout the next two phases of the study in order to generate the factors influencing the channel selection by older adults for these specific kinds of information needs.

Semi-structured interviews will be used in the second phase of the study in order to obtain a preliminary understanding of the factors affecting older adult's choice of information channels. These factors will be added to an expanded model of channel selection for older adults. This expanded model will be tested during the third phase of the study.

The third phase of the study will involve a survey to confirm the magnitudes and relationships between these factors. This survey will be pilot tested first for face and content validity with focus groups of older adults, and then pilot tested with small samples in order to determine construct validity and to estimate the sample size for the final survey [3]. Results from the final survey will be analyzed using Multiple Linear Regression to understand the impact of the various factors and their relationships [2].

The researcher hopes to use this information to build a comprehensive model of how various factors impact an older adult's choice of information channel, and in particular, the factors that affect if an older adult is likely to choose a digital versus a non-digital channel.

The poster presented at the conference will include preliminary results from the semi-structured interviews, including a revised model of the channel selection process detailed in Figure 1.

References

1. Blumler, J.G. The role of theory in uses and gratifications studies. *Communication Research, 6*. 9-36.
2. Cohen, J., Cohen, P., West, S.G. and Aiken, L.S. *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum, Mahwah, NJ, 2003.
3. Creswell, J.W. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage, Thousand Oaks, CA, 2003.
4. Czaja, S.J., Sharit, J. and Nair, S.N. Usability of the medicare health web site. *JAMA: Journal of the American Medical Association, 300* (7). 790-792.
5. Donohew, L., Palmgreen, P. and Rayburn, J.D. Social and psychological origins of media use: A lifestyle analysis. *Journal of Broadcasting and Electronic Media, 31* (255-278).
6. Hernández-Encuentra, E., Pousada, M. and Gómez-Zúñiga, B. ICT and older people: Beyond usability. *Educational Gerontology, 35* (3). 226-245.
7. Iyer, R. and Eastman, J.K. The elderly and their attitudes toward the internet: The impact on internet use, purchase, and comparison shopping. *Journal of Marketing Theory and Practice, 14* (1). 57-67.
8. Lam, J.C.Y. and Lee, M.K.O. Digital inclusiveness - Longitudinal study of internet adoption by older adults. *Journal of Management Information Systems, 22* (4). 177-206.
9. Laukkanen, T., Sinkkonen, S., Kivijarvi, M. and Laukkanen, P. Innovation resistance among mature consumers. *Journal of Consumer Marketing, 24* (7). 419-427.
10. Phang, C.W., Sutanto, J., Kankanhalli, A., Li, Y., Tan, b.C.Y. and Teo, H. Senior citizens' acceptance of information systems: A study in the context of e-government services. *IEEE Transactions on Engineering Management, 53* (4). 555-569.
11. Reisenwitz, T., Iyer, R., Kuhlmeier, D. and Eastman, J.K. The elderly's internet usage: an updated look. *Journal of Consumer Marketing, 24* (7). 406-418.
12. Roseman, G.H. and Stephenson, E.F. The effect of voting technology on voter turnout: Do computers scare the elderly? *Public Choice, 123*. 39-47.
13. Slegers, K., van Boxtel, M.P.J. and Jolles, J. The effects of computer training and internet usage on the use of everyday technology by older adults: A randomized controlled study. *Educational Gerontology, 33* (2). 91-110.
14. Tacken, M., Marcellini, F., Mollenkopf, H., Ruoppila, I. and Széman, Z. Use and acceptance of new technology by older people. Findings of the international MOBILATE survey: 'Enhancing mobility in later life'. *Gerontechnology, 3* (3). 126-137.
15. U.S. Census Bureau. Table 3. Percent Distribution of the Projected Population by Selected Age Groups and Sex for the United States: 2010 to 2050 (NP2008-T3). Division, P. ed., 2008.
16. U.S. Department of Health and Human Services. Medicare Modernization Update, 2003.
17. Umemuro, H. Computer attitudes, cognitive abilities, and technology usage among older Japanese adults. *Gerontechnology, 3* (2). 64-76.
18. Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27* (3). 425-478.
19. Wilson, T.D. Models in information behaviour research. *Journal of Documentation, 55* (3). 249-270.
20. Wright, D. and Hill, T. Prescription for trouble: medicare part d and patterns of computer and internet access among the elderly. *Journal of Aging & Social Policy, 21* (2). 172-186.
21. Zeithaml, V.A. and Gilly, M.C. Characteristics affecting the acceptance of retailing technologies: A comparison of elderly and nonelderly consumers. *Journal of Retailing, 63* (1). 49-68.

Categories and Subject Descriptors

H. Information Systems; H.1 MODELS AND PRINCIPLES; H.1.2 User/Machine Systems; *Human factors*

General Terms

Human Factors

Keywords

older adults, elderly, seniors, information seeking, media, digital information

Location-based questions and their implications for digital reference consortia

Bradley Wade Bishop
College of Information, Florida State University
101 Louis Shores Building
142 Collegiate Loop
Tallahassee, FL 32306-2100
bwb06c@fsu.edu

ABSTRACT
This poster addresses a lack of knowledge about location-based questions and the implications of this lack of knowledge on digital reference consortia. Location-based questions include any question that concerns the attributes of a georeferenceable location or locations. Findings indicate half of questions asked to the statewide service of this study were location-based questions. Recommendations from this study's findings include populating the consortium's knowledge base with local knowledge, especially participating information agencies local knowledge.

Categories and Subject Descriptors
H.3.5 [Online Information Services] web-based services.

H.3.4 [Systems and Software] question-answering (fact retrieval) systems

General Terms
Management, Human Factors.

Keywords
Digital reference, location-based questions, consortia.

1. INTRODUCTION
This poster presents findings from the content and quantitative analysis portions of a larger exploratory study, *Chat reference and location-based questions: A multi-method evaluation of a statewide chat reference consortium*. The content and quantitative analysis includes reviewing two months of transcripts from the Florida Electronic Library's *Ask-a-Librarian* service, 7,021 total transcripts. Chat reference and location-based questions refers to the question-negotiation process in the chat mode of responding to users' location-based questions.

In the statewide chat reference consortium of this exploratory study, users are able to ask any information provider from any of the one-hundred and three participating information agencies and any consortium information provider can respond to questions from any user. This situation may potentially lead to the assumption that because a local information provider is closer in proximity and more familiar to a location or locations within their same county, a local information provider may provide a higher correct response fill rate to location-based questions than a non-local information provider [1-5]. Although county boundaries may seem arbitrary, the boundaries have real world implications for governments to exercise control over others living within the boundaries, for example library users' service eligibility. Prior to assessing a correct response fill rate to location-based questions for local and non-local information providers in future studies, more exploratory study into the types of location-based questions was addressed.

2. LOCATION-BASED QUESTIONS
Location-based questions include any question that concerns the attributes of a georeferenceable location or locations. Types of location-based questions found in this study included directional questions (e.g. routes) and non-directional location-based questions (e.g. the question concerned attributes of a location or locations, including a point of interest, such as a library, and its circulation policies).

The non-directional location-based questions were subdivided into two major types of location-based questions, *geography* questions and *attribute of geography* questions. *Geography* location-based questions are about the location of a place (i.e., latitude and longitude) or concern the physical relation of a location to another location, as long as that place is not a library (e.g., where is Darfur?). As quantitative analysis will discuss, the majority of non-directional location-based questions concerned an attribute of geography of the location in a question rather than the location of a place or the physical relation of a location to other locations. These questions were named *attribute of geography* questions and concerned the attributes of a location or locations other than the attribute of physical location.

The researcher subdivided the attribute of geography questions into those concerning libraries, universities, and other locations. The study breaks out *library* and *university* location-based questions because these categories may lend themselves more to the provision of scripted responses, as they relate to the institutions participating in the chat reference consortium, as opposed to the other *attribute of geography* questions. The library and university location-based questions concern the attributes of a library or university. The other attribute of geography questions concern the attributes of an assortment of other places, ranging from Ancient Rome to Yosemite National Park.

3. RESEARCH QUESTIONS AND METHODS
The researcher utilized content and quantitative analysis to address its research questions.

1. What are the types of location-based questions?

2. What is the ratio of location-based question transcripts, in total and by type, to total transcripts?
3. What is the ratio of location-based question transcripts responded to by non-local information providers to total location-based question transcripts responded to?

Data collection and data analysis of content analysis provide qualitative data on the types of location-based questions. A grounded theory approach was used and question types emerged from a pilot study of from a ten-percent of the data. The researcher conducted all content analysis for consistency using these question types and because the work was a dissertation and therefore an independent study. To address issues related to intrarater reliability, the researcher coded 30 randomly selected transcripts using protocol from content analysis twice, allowing one month to pass between coding, in order to ensure intrarater reliability over time. An acceptable Cohen's kappa was 80 percent and the researcher obtained .860. To address interrater reliability, the researcher recruited and trained three external coders to code 30 randomly selected transcripts using protocol from content analysis. Coded material was compared across coders to ensure interrater reliability. An acceptable Krippendorff's alpha was 80 percent and the researcher obtained .8108. These statistics represented reliability for the entire protocol that included other elements beyond whether a question was location-based or not or the type of location-based question.

Measures of ratio from quantitative analysis provide quantitative data on the ratio of location-based question transcripts, in total and by type, to total question transcripts, and the ratio of non-local information providers responding to location-based question transcripts to total location-based question transcripts responded to.

4. FINDINGS AND IMPORTANCE

From a transcript population of two months of data from a statewide chat reference consortium (6,584), 50.2% of total transcripts were location-based question transcripts. The majority of location-based question types asked were library type questions. 78.3% of location-based question transcripts asked contained library location-based questions. In addition, 73.8% of information providers responding to location-based questions were non-local information providers.

By addressing these research questions, the findings provide some insight into the emerging information landscape of national and statewide chat reference consortia. The creation, adoption, and redefinition of information providers' service roles resulting from e-services, e-resources, and mobile devices frees some information providers and some users from the precondition of being proximally affixed to location-bound technologies and/or constrained by the operational hours of their information agencies [6, 7]. These same technological changes also broaden the geography from which both possible users with questions and potential information providers to offer responses may originate. The potential barriers for users and information providers participating in the statewide chat reference consortium may emerge from the content and quantitative analysis findings. The findings may also illuminate the preponderance of location-based questions and the issues related to response formulation by local and non-local information providers.

5. RECOMMENDATIONS

Recommendations from this study include populating the consortium's knowledge base with local knowledge, especially that from participating information agencies. The chat reference consortium manager may encourage participating information agencies to incorporate local knowledge commonly asked in location-based questions on their websites and into the services' knowledge base. Training information providers to use clarifying questions, to supply resources with responses, and to have a better understanding of the locations and policies of participating information agencies in the chat reference consortium as well as the geography of the state they serve may improve correct response fill rates. Chat software developers and chat consortia managers may mitigate some of the challenges associated with location-based questions by building geographic intelligence into their systems. This study indicates that chat consortia may overcome the potential weakness of location-based questions (i.e., referral), if participating information agencies improve their online dissemination of local knowledge that are related to questions users actually ask.

6. REFERENCES

[1] Berry, T. U., Casado, M. M., and Dixon, L. S. 2003. The local nature of digital reference. *The Southeastern Librarian, 51*(3) (Fall): 8-15.

[2] Bishop, B.W. and Torrence, M. 2007. Virtual reference services: Consortium vs. stand-alone. *College and Undergraduate Libraries, 13*(4): 117-127.

[3] Hyde, L. and Tucker-Raymond, C. 2006. Benchmarking librarian performance in chat reference. *The Reference Librarian, 46*(95/96): 5-19.

[4] Kwon, N. 2007. Public library patrons' use of collaborative chat reference service: The effectiveness of question answering by question type. *Library & Information Science Research*, March 2007, 29(1): 70-91.

[5] Sears, J. 2001. Chat reference service: An analysis of one semester's data. Issues in Science & Technology Librarianship. Retrieved June 11, 2008 from http://www.istl.org/01-fall/article2.html.

[6] Morville, P. 2005. *Ambient Findability*. Sebastopol, CA: O'Reilly Media, Inc.

[7] McClure, C. R. and Jaeger, P. T. 2008. *Public libraries and Internet service roles: Measuring and maximizing Internet services*. Chicago, IL: American Library Association.

Assessing Need for an Automated File Format Obsolescence Warning System for Digital Collections

Heather L.M. Bowden
University of North Carolina at Chapel Hill
216 Lenoir Drive
CB#3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
hbowden@email.unc.edu

ABSTRACT

Anecdotal evidence reported in literature, and personal discussions with managers of digital archives suggests that one of the greatest hindrances to the successful preservation of resources in digital archives is the high level of repeatable activities that are required to be performed in order to monitor their digital collections' viability over time [1] [2] [3]. Equally troublesome is the rate at which digital file formats become "obsolete," or not readable by current computer software and/or hardware. It is not currently clear which tools should be developed to best ameliorate these issues, or the severity of the actual needs for these types of tools in the digital archives environment.

At present there is no fully functioning system which can detect and notify digital archives managers of impending file format obsolescence. In order for preservation systems to evolve and grow in step with the changing technological landscape, they need to find a way to dynamically monitor and react, if necessary, to the changes as they occur. Static systems with rigid controls of data flow have no ability to monitor, adapt, and grow as the sands of technology shift. 'Community watch and participation' is a key component of the DCC Curation Lifecycle [5], but has yet to be formally applied to functions in developing preservation systems.

In order to begin designing tools which will aid in the management and preservation of digital collections, the first step is to engage with the community of digital collection managers and learn directly from them about their needs in this arena. Using the principles of user centered design, the following study was conducted as a first step in the iterative design process to create an automated file format obsolescence warning system. This is part of the initial design phase of "collecting critical information about users," [4] which will lead to the iterative cycle of design, test and measure, and redesign.

This study seeks to answer three research questions: 1) What types of file formats are currently being managed in digital collections, 2) What methods are digital collection managers currently employing to sustain their collection over time, and 3) What types of tools (automated or otherwise) can help digital collection managers in sustaining their collection over time?

The information collected from this study will be used to inform the development of a file format obsolescence warning system which will make use of collective intelligence and community participation in order to dynamically monitor and report on the changes in file format viability.

Data was collected for this study through semi-structured phone interviews with managers of digital collections; and have been qualitatively analyzed using grid analysis techniques in order to assess patterns, consensus, and outlier information about their collections, preservation practices, and needs for tools in managing file format obsolescence.

A total of nine participants took part in this study and are all professionals who are responsible for the management of a digital collection. They all answered questions about the digital collections they managed. These questions were broken down into six broad categories: 1. Which file formats are you currently managing?, 2. For how long are you intending to or required to preserve the digital items in your collection?, 3. What aspects of your digital collection are most important to preserve?, 4. What measures do you take or activities do you currently perform to manage file format obsolescence in your collections?, 5. Would an automated file format obsolescence notification system be helpful?, and 6. What other tools could help you?

The following generalized answers to these questions are being applied to further research and tool development. The range of file formats managed across collections varied widely, where the most common file formats (TIFF and PDF), were found in almost all of the collections. The respondents were most concerned about preserving the more obscure file formats such as DBASE and Déjà Vu. Every collection manager reported that the items in their collections were expected to be preserved indefinitely. Each digital collection specified different properties of the digital objects which needed to be preserved. Even in the same collection, there were different properties which were important to preserve in different contexts. There was a wide range of digital preservation activities being performed across the collections, from "nothing " to "educate the data producers" and the implementation of a migration on ingest program. Where every participant responded affirmatively that they could benefit from having an automatic file format obsolescence notification system, they all had different visions of how it

could be implemented in their workflow. Other tools which were reported to be desired were automatic validation & authenticity checking functions and automatic migration functions.

Implications of the study results point to the need to develop an automatic file format obsolescence/endangerment notification system which can assess a wide range of file formats for an indefinite period of time. The system must also allow for granular user controls which can be implemented not only at the institutional level, but also at a use case levels. Most importantly, any such system must be able to evolve and change in step with the technological landscape it is monitoring.

A prototype of a system which will address these needs will begin to be developed in the summer of 2010. A proposed, high-level conceptualization of this system is shown in *Figure 1*. In this model, a technology watch component is comprised of a collective intelligence unit and sorting and analyzing algorithms which work together to create the output of a list of file formats and their endangerment warning levels.

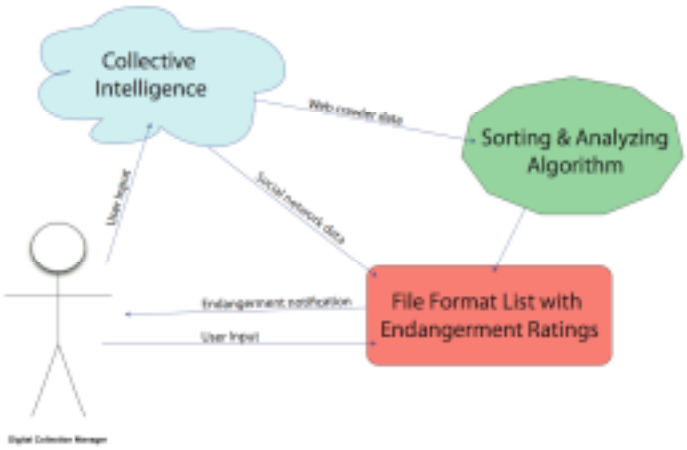


Figure 1. Basic model of the proposed system structure .

The collective intelligence component is comprised of data pulled or "crawled" from websites as well as data informed by a combination of loose social networks and tight, predetermined social networks. Collective intelligence has been generally defined as, "when a group of individuals collaborate or compete with each other, intelligence or behavior that otherwise didn't exist suddenly emerges." [6] When referring to technology, it has also been said to be the "combining of behavior, preferences, or ideas of a group of people who create novel insights." [7]

The sorting and analyzing algorithms are based on the CUSUM algorithms used by the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) system, which are used to analyze data collected from several sources online in order to detect outbreaks of infectious diseases. [8]

The output of these two components is the dynamically generated and updated list of file formats and their endangerment ratings. Input from the pre-determined social networks is used to refine the list to the specific needs of the group and input from the individual digital collection managers is used to refine the list further for the needs of their institution and individual use cases. The individual digital collection manager may also inform the system less directly by sharing their knowledge and experience via any channel on the World Wide Web.

By using collective intelligence methods and models developed for other early warning systems, it will be possible to provide more timely and relevant file format endangerment warnings to digital collection managers. This system design allows for the inclusion of all file formats and also allows for specifications to be changed on an individual and context specific level. The information collected in the participant interviews shows that these capabilities are relevant to their needs and important in the design of a file format endangerment warning system, and so they have been incorporated into the first design stage of this project. Further research and user testing will be conducted as test systems are implemented.

1. REFERENCES

[1] Brown, A. 2005, January 4. "PRONOM 4 Information Model, version 1", United Kingdom: The National Archives.

[2] Giaretta, D. 2007. "The CASPAR approach to digital preservation", The International Journal of Digital Preservation, 1, 2 , 112-121.

[3] Hedstrom, M. 1998. "Digital preservation: A time bomb for digital libraries", Computers and the Humanities, 31, 189-202.

[4] Gould, J.D., & Lewis, C. 1985. "Designing for usability: Key principles and what designers think", Communications of the ACM, 28, 3, 300-311.

[5] Higgins, S. 2008. "The DCC Curation Lifecycle Model", International Journal of Digital Curation, 3, 1, 134-140.

[6] Alag, S. 2009. "Collective intelligence in action" Greenwich, CT: Manning Publications.

[7] Segaran, T. 2009. "Programming collective intelligence", Sebastopol, CA: O'Reilly Media.

[8] North Carolina Division of Public Health (NC DPH). Retrieved May 5 2009 from http://www.ncdetect.org.

Improving Federal Policy on Website Accessibility

John Brobst
Florida State University
1048 Kingdom Drive
Tallahassee, FL 32311
(850) 562-8585
johnbrobst2000@yahoo.com

ABSTRACT

This paper provides an abstract for a proposed poster. The poster will present preliminary findings of research conducted as part of a doctoral dissertation effort.

This paper conducts a policy analysis of the current United States federal policy on website accessibility. Website accessibility means "making the web useable by everyone whatever their ability or disability" [1]. Previously, web accessibility meant the providing of equal access and equal opportunity to the Internet for people with disabilities. More specifically, Web accessibility means that people with disabilities can perceive, understand, navigate, and interact with the Web, and that they can contribute to the Web.

The problem this study addresses is to determine how well federal health care websites comply with Section 508 of the Rehabilitation Act of 1973 (29 U.S.C. §794d.). The purpose of this research study is to improve accessibility of federal (United States government) healthcare websites. The intent of this study is to develop a better understanding of the accessibility of federal health care websites and to determine what information policy barriers may prevent attaining full accessibility.

Research Questions:

- Are federal health care websites compliant with existing legislation on website accessibility?
- Do current laws articulate a clear and consistent federal policy on web accessibility?

This study is a multi-method approach which includes a literature review plus two methods of policy analysis. The first method is a case study of a distinct category of federal websites. This examination provides an indication as to the degree of success attained by the legislative acts that intended to establish website accessibility.

The second method of evaluation is a classical policy analysis approach that performed a side-by-side analysis of the two legislative acts that most closely relate to federal web accessibility policy. The information policies examined are the Americans with Disabilities Act of 1990 and the Rehabilitation Act of 1973, as amended by Section 508 in 1990.

The Technology Acceptance Model (TAM) has been used extensively in information technology related areas, and has been successfully used to as a conceptual framework supporting web accessibility research [2][3]. The TAM provides a framework for understanding how users adopt and use new technologies [3][4]. TAM lends itself to the study of web accessibility as it captures its innate complexity, as it is a topic that contains aspects of policy issues, technical considerations, and integrates issues that are user centric [2]. Jaeger studied federal website accessibility, selecting a modified TAM model as the most relevant model to explore as an appropriate conceptual framework for this type of investigation. Jaeger offered observations about the value to the modified TAM as a conceptual framework for web accessibility applications. Jaeger's study indicated the possible need to incorporate four more factors or influences into the modified TAM framework. Those factors were: user feedback; education and training; monitoring and enforcement; and political climate. The research performed in this study will determine the appropriateness of including those factors in the TAM model for application to web accessibility research. This study could extend the TAM framework by addressing three of the four areas identified by Jaeger's work: education and training; monitoring and enforcement; and political climate. This study will assess the value of those areas towards the incorporation of those three factors into the modified TAM conceptual framework, with an intended outcome of further developing and improving the TAM model for use in conducting theory based web accessibility research.

The case study of a distinct category of federal websites provided an indication as to the degree of success attained by the legislative acts that intended to establish website accessibility. The case study examined all the healthcare related websites provided by the official federal government portal website (WWW.HEALTHCARE.GOV). Of the 35 website homepages examined, 8 had accessibility errors. The finding is that 23% of these websites failed to be fully accessible as mandated by current web accessibility legislation. This evaluation indicates that the intent of federal web accessibility policy has not been fully achieved.

The side-by-side analysis of the legislative acts (Americans with Disabilities Act of 1990 and Section 508 of the Rehabilitation Act of 1973) revealed significant differences in the ways these acts attempt to eliminate discrimination

against individuals with disabilities. The most significant difference was in how these acts portrayed the intended recipient of these civil protections. The Rehabilitation Act perspective was that of *accommodating* an "*individual with a handicap,*" versus the ADA perspective of providing *universal accessibility* to "*individuals with a disability*. This evaluative comparison helped to reveal the strengths and weaknesses of each act, and provided indication for improving these legislative efforts towards attaining web accessibility.

The interim findings from this study led to the development of five policy options that could serve to significantly improve the effectiveness of the federal information policy regarding website accessibility. Assessing the merits of these options involved the use of several evaluative criteria: effectiveness, feasibility, cost, and political impacts.

That evaluation produced a recommendation for pursuing two interrelated options:

- Requiring all federal website managers to receive mandatory training leading a certification in web accessibility.
- Committing to ongoing research efforts to determine the impediments to attaining web accessibility and to identify the best practices that promote attaining web accessibility.

These options are interrelated as the knowledge gained from the ongoing research effort would serve to as feedback loop into the certification training programs. This feedback provides for continuous improvement in the skills, knowledge, and abilities of the federal web managers. These recommended policy options would help the federal government to better comply with the intent of the existing legislation and to assure fully accessible federal government websites. While the federal government has been progressing toward equal participation in its government for individuals with disabilities for 3 decades, the federal government must now focus on assuring that all individuals can have fair and equal participation in the new frontier of cyberspace.

A significant value of this research lies in its uniqueness. This study explores a relative uncharted area, being the accessibility of health care websites provided by the United Stated federal government. In the United States, federal websites are required to be accessible as defined by the criteria identified in Section 508 of the Rehabilitation Act of 1973. Therefore, the accessibility of government provided health care websites is mandated by this legislation. The expectation of the law is to assure that a large segment of society (those individuals with disabilities) are not effectively excluded from the benefits, services, and products offered by these websites [5]. An inaccessible federal health care website would effectively deny persons with disabilities the chance to use the information and services in a fair and equal manner [2].

From a policy perspective, this research is important in its potential to contribute to improving the policies and legislation that attempt to attain the social goals of equity and fairness for all citizens of this nation. By examining the key policies that relate to federal health care websites, this research may reveal areas of conflict or inconsistencies that act as impediments. This study looks for those inconsistencies or issues by examining the key legislative documents using a classical policy analysis approach, and will be looking for policy issues by surveying federal web site managers. By searching for policy conflicts and issues from these two approaches, it is expected that a richer and fuller understanding will be achieved. This type of analysis can help inform the legislative process, and may indicate whether the social goals of providing an accessible federal government are being realized in the manner prescribed by existing legislation.

This research will likely be of greatest importance and of direct value to the over 54 million Americans with disabilities, as it may improve their ability to access the information and services that are provided by federal health care websites [6]. The intent of this research effort is to facilitate increased accessibility of federal health care websites. In doing so, individuals with disabilities will be better able to use, comprehend, and interact with the content, services and products that are offered to the American public through these federal websites.

The growing importance of providing information using Internet based systems serves to support the need to study this area. The intended outcome of this research is to improve accessibility levels and assure that equal access exists for all individuals as a matter of social justice. Assuring web site accessibility serves to guarantee that individuals with disabilities will have a level of use that is equal to the use enjoyed by those individuals that do not have such disabilities.

General Terms

Legal Aspects.

Keywords

Website accessibility, information policy, Section 508.

REFERENCES

[1] W3C. World Wide Web Consortium .2009. (W3C). Retrieved on April 7, 2008, from http://www.w3.org/

[2] Jaeger, P. 2006. Assessing Section 508 compliance on federal e-government websites: A multi-method, user-centered evaluation of accessibility for persons with disabilities. Government Information Quarterly, 23(2), 169-190.

[3] Davis, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly. 13, 319-339.

[4] Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. 1989. User acceptance of computer technology: A comparison of two theoretical models. Management Science. 35, 983-1003.

[5] Access Board. 1999. (dated May 12). Electronic and Information Technology Access Advisory Committee - Final Report. Retrieved July 24, 2009, from http://www.access-board.gov/sec508/commrept/eitaacrpt.htm

[6] Pew Internet and American Life Project. (2004). P*eople with disabilities.* Washington, DC. Retrieved March 13, 2008, from http://www.pewinternet.org/reports/

PREPARING FUTURE DIGITAL CURATION FACULTY: THREE DOCTORAL FELLOWS AS EXAMPLES

Sarah Ramdeen, Kaitlin L. Costello, Mike Brown

School of Information & Library Science
University of North Carolina
Chapel Hill, NC USA
+1 919-962-8366
ramdeen@email.unc.edu, kaitcost@email.unc.edu, brownstudy@email.unc.edu

ABSTRACT

Digital curation, the curation of digital assets, whether cultural, educational, scientific, or economic, is emerging as an active field of research and development.

"Digital curation" can be defined as "the active management and preservation of digital resources over the life-cycle of scholarly and scientific interest, and over time for current and future generations of users"[1]. It involves "maintaining and adding value to a trusted body of digital information for current and future use;" [2] and is "key to reproducibility and re-use"[1]. Limited graduate educational opportunities in digital curation exist. The School of Information and Library Science (SILS) at the University of North Carolina at Chapel Hill (UNC-CH) was awarded a 2008 Laura Bush 21st Century Librarian Program Grant from IMLS, under the category "Programs to Build Institutional Capacity." DigCCurr II: Extending an International Digital Curation Curriculum to Doctoral Students and Practitioners" (2008-2012) builds upon DigCCurr I (2006-09). SILS continues in their partnership with NARA, as well as introducing a new partnership, the University of Glasgow's HATII. This project seeks to develop a doctoral-level curricular framework; course content; and networked, distributed, international seminars to prepare future faculty to educate 21st century digital curators.

The DigCCurr II project will also offer three multi-stage Institutes for Professionals in Curation Practices for the Digital Object Lifecycle, taught by a team of international experts, designed to have an immediate impact on curation practices. The Institutes are designed to foster skills, knowledge and community-building among professionals responsible for the curation of digital materials. A highly interactive website, the Digital Curation Exchange [3], has been developed which acts as a forum for discussion and collaboration during the institute and also provides forums for discussion, sharing of documents, as well as links to valuable tools and resources.

The DigCCurr II Program also provides funding for six Carolina Digital Curation Doctoral Fellows. The doctoral fellows are assisting the Institute instructors in course development and implementation through out the year and during the Institute. Each Fellow also brings his/her own knowledge and specialties to the institute, and are also offered the opportunity to develop their individual research interests and skills in addition to gaining valuable teaching experience the work of the fellows will help enrich the collective knowledge of digital curation practices. There are currently three doctoral fellows. This cohort will be expanded to include three additional Fellows who will be recruited for Fall 2010.

This poster will report on the research trajectories of the Carolina Digital Curation Fellows as an illustration of how doctoral-level education is being integrated into a digital curation curriculum.

Mike Brown is currently working on a project studying job market ads related to digital preservation and curation. He has specific interests in the digital curation of cultural heritage and performing arts artifacts, and the overlap between personal information management and personal archiving strategies.

Kaitlin Costello's research areas include digital curation and digital archives. Specifically, she studies selection and appraisal methods for interactive digital objects. She is also interested in instructional design and pedagogy. Kaitlin is currently conducting a study examining digital preservation education at the graduate level in information schools. She is also researching appraisal strategies for websites created by UNC.

The Sarah Ramdeen is interested in digital curation and education for the sciences – teaching others to manage their digital collections and shaping the way scientists preserve and access their data. Her current research areas include a qualitative study of collection managers at geological repositories and she is currently developing a study to look at the preservation and archival habits of geological researchers and students in relation to specialized file formats.

DigCCurr II seeks sustainability for work in the digital curation arena through raising public and professional awareness. By cultivating intensive deliberation and mutual engagement of the issues among a diversity of players, DigCCurr II will also contribute substantially to the promotion and sustainability of the communities of practice that will ensure responsible, long-term digital curation.

While all of the Fellows view digital curation through a unique lens, their projects are integral to the development of their skills as researchers. Their work will bring awareness of the critical issues of the field and will enhance the current body of knowledge in the practice of digital curation. Through their experiences with the Institutes and their collaboration with international experts, they will be better prepared for careers as educators of the next generation of digital curation practitioners.

Categories and Subject Descriptors

K.3.2 [**Computers and Education**]: Computer and Information Science Education –/ curriculum.

Keywords

Digital preservation, graduate education, digital curation

1. ACKNOWLEDGMENTS

This work was supported through IMLS Grant #RE-05-08-0060-08. We are grateful to our anonymous reviewer and Dr. Cal Lee and Dr. Helen Tibbo for their thoughtful comments.

1. REFERENCES

[1] Digital Curation Centre, “What is Digital Curation?” http://www.dcc.ac.uk/about/what/

[2] Digital Curation Centre, “About the DCC.” http://www.dcc.ac.uk/about/.

[3] Digital Curation Exchange, http://digitalcurationexchange.org/

Virtual Scientific Teams: Life-Cycle Formation and Long-Term Scientific Collaboration

Gary Burnett
Kathleen Burnett
Michelle M. Kazmer
Paul F. Marty
Besiki Stvilia
Charles C. Hinnant
Adam Worrall
Florida State University College of Communication & Information
PO Box 3062100
Tallahassee, FL 32306-2100
1 850 644 5775

ABSTRACT
Researchers will model the lifecycles of virtual multidisciplinary scientific teams using the facilities of the National High Magnetic Field Laboratory, an interdisciplinary scientific center with distributed facilities in Tallahassee, Florida; Gainesville, Florida; and Los Alamos, New Mexico. The model will be built from data collected through descriptive multiple-case studies, grounded in an analysis of social and organizational factors related to the concepts of the theory of information worlds: social norms, social types, information values, and information behaviors (Burnett & Jaeger, 2008; Jaeger & Burnett, in press). The researchers hypothesize that when the norms and practices of multiple external worlds represented by team members are integrated into the internal norms and practices of the team itself, the outcomes of the project will more likely be successful, and team members will be more likely to work together virtually again.

Categories and Subject Descriptors
K.42.3 [**Computers & Society**]: Social issues – *employment.*

General Terms
Human Factors, Theory.

Keywords
Scientific collaboration, life cycle models, virtual organizations.

1. INTRODUCTION
There are increasing efforts to build an advanced infrastructure for e-science, including high performance computing centers, connected through high speed networks, to facilitate the sharing of both instruments and datasets, and to enable more effective scientific collaborations, learning and professional development (Atkins, Droegemeier, Feldman, Garcia-Molina, Klein, Messerschmitt, Messina, Ostriker, & Wright, 2003).

However, neither infrastructure nor applications guarantee that scientists will use the technology, establish successful collaborations, or share data. Cultural and social factors may either constrain or encourage the adoption and use of technology or data. Similarly, the technology may influence social structures and enable or constrain social interaction, data sharing, and collaboration (Birnholtz & Bietz, 2003; Orlikowski, 1992; Stvilia, Twidale, Smith, & Gasser, 2008).

To improve understanding of the sociotechnical factors affecting lifecycle development, this research asks *what social and organizational factors best support the transition of short-term experiment-focused multidisciplinary virtual scientific collaborations to long-term productive and innovative programs of scientific research?* The goal is to develop and validate a lifecycle model to support distributed scientific teams through the transition from discrete experiment-focused projects to long-term distributed collaborations, thereby advancing innovation and increasing productivity.

2. RESEARCH DESIGN
The project draws its framework from the *theory of information worlds*, which seeks to describe intertwined information exchange and social interaction in a variety of settings. The information worlds of the short-term scientific teams under investigation are *intrinsically transient*, with pre-defined ending points, after which they will cease to exist; thus, they exhibit distinct lifecycles (including specific beginning and ending points). The nature and specifics of the teams' lifecycles have important implications for their interactions, for how they exchange information, and for their success or failure.

One sub-set of research questions seeks to determine how virtual organizations demonstrate that they perform successfully. The specific dimensions along which performance is assessed include willingness to work together again with the same colleagues; willingness to work on virtual teams again; and research output:

1. *Is there evidence that the lifecycle of a virtual team influences the willingness of individual team members*

to work together again? How does this compare with their willingness to work together again with co-located team members?

2. *Is there evidence that the lifecycle of a virtual team influences the willingness of individual team members to work in virtual teams again? How does this compare with their willingness to work in co-located teams again?*

3. *Do virtual teams generate output as measured by patents, journal articles, and presentations comparable to the output of co-located teams working on similar projects? Is there a difference in the amount of time required to generate such outputs?*

One question investigates the multi- and interdisciplinary research collaborations fostered by the Magnet Lab:

4. *Is there evidence to suggest that the degree of multi- or interdisciplinarity within a team influences its lifecycle or its outcomes?*

The study will also examine relationships between types of teams and team performance:

5. *Do collaborating groups share a definable set of norms and expectations regarding how CMC-based interactions are supposed to function in order to ensure successful collaborations?*

6. *If there are such norms, do they appear to be established ad-hoc by the collaborating groups, or are they established (formally or informally) externally to the groups, and adopted as part of the working strategies of the groups?*

7. *Is there evidence of conflicting norms, or of multiple "information worlds" coming into contact or conflict during collaborations?*

8. *Is there evidence of different types of virtual teams and projects in the research sample, particularly in terms of the different external worlds represented by team members, and are such differences linked to team outcomes or to team members' willingness to work in virtual teams again?*

3. DATA COLLECTION AND ANALYSIS

The research will collect data through descriptive multiple-case studies, and will include content analysis and social network analysis based on observations, interviews, and documentary artifacts of virtual communication generated by scientific teams.

Researchers will collect and analyze a wide variety of documentary artifacts, beginning with a convenience sample of artifacts representative of interactions of teams who completed projects in 2009. Collection and analysis of this sample will provide an opportunity to test assumptions about the nature and purpose of interaction and to develop preliminary classifications along several dimensions. Modes of interaction may be synchronous or asynchronous, and relationships may be one-to-one, one-to-many, or many-to-many. Media may include print, audio, visual, or audiovisual.

Direct observations will be conducted of the multidisciplinary teams selected for the multiple-case study. Observations will occur at the Magnet Lab while teams are conducting their experiments, typically over the course of one week.

Members of each scientific team will be interviewed following the completion of their experiments and researchers' analysis of the documentary artifacts and observation reports. Factual incidents will be collected from the documentary artifacts and direct observations. Themes will be identified, the incidents will be sorted into categories, and questions developed.

This study will collect textual and audio data, which will be analyzed using Nvivo software. Although it will be necessary to transcribe audio data, Nvivo allows it to be tagged and stored for ease of retrieval, allowing researchers to recall the data in context. Analysis will employ a codebook including codes for social norms, social types, information value, information behavior, project types, and lifecycle phases. It will also employ in vivo coding, allowing codes to be assigned directly from the audio or textual utterances as they are originally portrayed. In vivo coding ensures that unexpected findings will not be overlooked.

Coding will be compared to ensure intercoder reliability, and where there are inconsistencies, the researchers as a group will make decisions that can be incorporated into subsequent coding. Analysis and interpretation of the data will be the responsibility of the researchers, who have extensive experience in conducting content analysis.

The techniques of social network analysis (SNA) (Wasserman & Faust, 1994) have been widely used to explore group structure, and test hypotheses about dynamics, interaction, information flow, knowledge acquisition, and diffusion. This project will combine the techniques of SNA and content analysis with data and text mining to supplement the qualitative analysis of virtual team behavior with additional information about teams' structural properties and relationships. The Magnet Lab will provide the researchers with access to the electronic communication logs. The researchers will also have access to the Magnet Lab's report and publication repository (http://www.magnet.fsu.edu/usershub/publications/index.html). These data sources, together with interview data and patent information, will be used to construct team social networks and mine for relationships. The study will use Pajek and Stata software for network analysis, visualization, and statistical relationship testing. The social network and statistical analyses tools will be used to identify and test the relationship between structural measures (e.g., network path length, density, centralization, clustering coefficient) and team and project types; model team dynamics; test the relationships between structural characteristics, productivity, and team type (i.e. virtual, co-located) and productivity; and likelihood of scientists joining the team.

4. CONCLUSION

The lifecycle model(s) developed in the research will enable multidisciplinary virtual scientific teams to better exploit computer-mediated communication technologies to extend their lifecycles from discrete projects to the long-term programs of research required to solve complex scientific problems. Every effort, including external evaluation, will be made to ensure that the model may be generalizable to other federally funded national laboratories, as well as to private sector scientific collaborations, thus enhancing national scientific productivity and global competitiveness.

The model(s) are expected to contribute to the advancement of both practical and theoretical knowledge: 1) within the domain of collaborative scientific inquiry, the model(s) will enable virtual multidisciplinary scientific teams to better exploit computer-mediated communication technologies to extend their lifecycles from discrete projects to the long-term programs of research required to solve complex scientific problems; 2) within the domains of social informatics and the science & technology studies, the model(s) will provide a framework for implementing theoretically-informed future research on virtual organizations and sociotechnical systems.

5. REFERENCES

[1] Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., Messina, P., Ostriker, J. P., & Wright, M. H. (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. Retrieved May 25, 2008, from http://www.nsf.gov/oc/oci/reports/atkins.pdf

[2] Birnholtz, J. , & Bietz, M. (2003). Data at work: supporting sharing in science and engineering. In: *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*, 339-348.

[3] Burnett, G. & Jaeger, P. T. (2008) Small worlds, lifeworlds, and information: The ramifications of the information behaviors of social groups in public policy and the public sphere. *Information Research*, 13(2). Retrieved May 11, 2009, from: http://InformationR.net/ir/13-2/paper346.html

[4] Jaeger, P.T., and Burnett, G. (In press). *Information Worlds: Social Context, Technology, & Information Behavior in the Age of the Internet*. Routledge.

[5] Orlikowski, W. (1992). *Learning from notes: organizational issues in groupware implementation.* In: Proceedings of the Conference on Computer-Supported Cooperative Work. Toronto, Canada, 362-369.

[6] Stvilia, B., Twidale, M., Smith, L. C., & Gasser, L. (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology, 59*(6), 983–1001.

The Study of Information Revisited: Chaos in the Emergence of Disciplinary Identity

Kathleen Burnett
Florida State University College of Communication & Information
PO Box 3062100
Tallahassee, FL 32306-2100
1 850 644 8124
kburnett@fsu.edu

Laurie J. Bonnici
University of Alabama
College of Communication & Information Sciences
Box 870252
Tuscaloosa, AL 35487-0252
1 205 348 8824
lbonnici@slis.ua.edu

Manimegalai M Subramaniam
College of Information Studies
University of Maryland
2118F Hornbake Bldg, South Wing
College Park, MD 20742
1 301 405 3406
mmsubram@umd.edu

ABSTRACT
This research will develop time-series fractal maps of LIS and CIS. The maps will be used to trace the trajectory of Information Science from its beginnings in 1965 into the future. Following Machlup and Mansfield (1983), we plan to analyze the logical, methodological, and pragmatic relations among and between these two areas of study centered on information creation, access, distribution and use. Our goal is to facilitate understanding of the future trajectories of the discipline of Information Science, and thereby, those of LIS and CIS through systematic examination of past and present trends. Because these trends are neither linear in their progression nor take place in a vacuum, our analysis will be guided by the fractal theory developed by Andrew Abbott in *Chaos of Disciplines*.

Categories and Subject Descriptors
D.3.3 [**Computers & Society**]: Social issues – *miscellaneous*.

General Terms
Human Factors, Theory.

Keywords
Information science, disciplinary identity, history.

1. INTRODUCTION

In *The Study of Information: Interdisciplinary Messages*, Machlup and Mansfield (1983) identified more than forty disciplines and sub-disciplines engaged in the study of information between the end of the Second World War and the publication of the volume. They invited representatives of a variety of disciplines to "analyze the logical (or methodological) and pragmatic relations among the disciplines and subject areas that are centered on information" (p. 3). Included were sections on two different—although perhaps overlapping—disciplines bearing the name Information Science:

Computer and Information Science (CIS), and Library and Information Science (LIS), each having its origins in the early sixties. The authors of the two sections separately point to a common semantic quirk that sets these disciplines apart from the rest: each is coupled or "anded" to another "science."

More than a quarter of a century later, the scholarly discourse generated by the book retains its vitality and interest. While the titles associated with the two disciplines have persisted, variations have been tested and discarded. Informatics, suggested by Gorn as an alternative to Computer and Information Science, had short-lived currency in each of the disciplines. Many of the academic units that house LIS programs have removed "Library" from their names, but their programs continue to focus, albeit to varying degrees, on libraries as institutions and librarianship as practice. Most recently, a relatively small interdisciplinary group (twenty-two institutional members as of April 2009) has announced its intention to form a new "iField." Does this announcement herald a convergence, or will the iField become the next battleground in a seemingly unending turf war?

2. RESEARCH DESIGN

Our analysis will begin with data collected from a wide variety of sources using a mixed methods approach. Abbott's analytic framework will be applied to clarify the nature of the evolution of these disciplines as Traditional differentiation, Fractal differentiation, or Fractal cycle. We will summarize and examine the outcomes of the research to determine how the information disciplines have evolved over the past 45 years with an eye to predicting their future trajectories.

The analytic framework will provide the predictive power needed to suggest whether the iSchools' iField initiative will culminate in convergence or in an extension of the ongoing turf war. We will combine content analysis, interviews, co-citation network analysis, MPACT metrics, and information visualization techniques to develop ten "snapshot" maps at five-year intervals to evoke the shifting terrain and identify the formation, dissolution and reformation of invisible colleges. The fractal

maps of the disciplines of LIS and CIS for the period from 1965-2010 will be compared to the descriptions and trajectories suggested by the experts contributing to the two sections in Machlup and Mansfield (1983).

Based on a review of the literature we expect three distinct eras to emerge from the analysis: *Conflict & Ingestion: The Development of Information Science, 1965-1980*; *Mainframe/ Library to PC: The Evolving Role of the Human Intermediary in Information Access, 1980-1995*; and *Place/Library vs. Virtual Space/Digital Libraries: The Internet and Disciplinary Identity, 1995-2010*.

3. ERAS

3.1 Conflict & Ingestion: The Development of Information Science, 1965-1980

The year 1965 was selected to begin the analysis since it is usually recognized as the date when the term "information science" first became associated with the two academic disciplines under discussion. In that year, the American Documentation Institute changed its name to the American Society for Information Science. Taylor's "Professional aspects of information science and technology," published that year, includes the earliest definition of the term.

Using the analysis techniques outlined above, we will to develop four snapshot maps at five-year intervals (1965, 1970, 1975 and 1980) to evoke the shifting terrain and identify the formation, dissolution, and reformation of invisible colleges during this formative period. The results will be used to evaluate the predictions and conclusions drawn by the authors contributing to relevant chapters of The Study of Information, and will be concatenated with those for subsequent eras for analysis of trends over time.

3.2 Mainframe/Library to PC: The Evolving Role of the Human Intermediary in Information Access 1980-1995

The human role in the information access equation is expected to emerge as the defining characteristic of the period from 1980-1995, particularly as it relates to organizational change within the disciplines of LIS and CIS. Humans, processes, and context comprise organizations. Organizations reorganize and redefine themselves as a result of external pressure. Forces of external pressure include technological advancements, competition, economics, and politics. Cycles of change grow shorter as a result of the increase in developments in information technologies and the social and economic contexts of information. The advent of the Internet, particularly the World Wide Web, fueled a paradigmatic shift in the information production and access environments driving the information science professions into discontinuous change. In order to determine which of the mechanisms: traditional differentiation, fractal differentiation, or fractal cycle is operating in the evolution of the LIS and CIS fields toward an iField, we undertake a macro-analytical approach to examining changes in labels and descriptions of the organizations that educate information professionals in both disciplines and the research foci of their respective faculties.

To capture shifts in time, analysis will be partitioned into 3 5-year segments; 1980-1985, 1985-1990, and 1990-1995. Interviews will be conducted with the three surviving founding deans of the iSchools movement to provide historical perspectives on the forming of the iSchools consortium. Content analysis of archived and current school and professional organization websites, postings to the Jesse listserve, scope statements of LIS and CIS research journals, and abstracts and papers from professional conferences the researchers will be conducted to identify how human decisions have shaped organizations. Changes in school names, repositioning of organizational philosophy as evidenced through restatement of organizational missions, shifts in faculty expertise as expressed through job vacancy announcements, and evolution of ideas through research method and scope will be examined. Of particular interest is the intersection or differentiation of the concept of information as defined by the LIS and CIS disciplines in the context of the iField movement.

3.3 Place/Library vs. Virtual Space/Digital Libraries: The Internet and Disciplinary Identity, 1995-2010

For the period from 1995-2010, we will continue the analysis of information science by examining the social and technical implications of the Internet on the development of disciplinary identity of LIS and CIS. The year 1995 was chosen because it marks the beginning of the "dot.com bubble," during which stock market values in many Western nations, and particularly in the US, increased rapidly from growth in the new Internet sector and related fields. The emergence of the Internet and the World Wide Web enabled expanded applications in these dotcom businesses, allowed remote presence, invoked the existence of digital libraries, and brought upon more challenging roles for databases,

human computer interaction and integrative programming. In addition, 1995 marks the emergence of information technology degree programs within LIS and CIS. More recent changes, such as the emergence of Web 2.0 and social networking tools further complicate the disciplinary identity formation process for both CIS and LIS.

Three fractal maps of the disciplines of CIS and LIS will be developed for the period from 1995- 2010, to illustrate the changes in disciplinary identity and compare these to those of the earlier periods. To depict change over time, analysis will be partitioned into three 5-year segments; 1995-2000, 2000-2005, and 2005-2010. We will determine which of the mechanisms: traditional differentiation, fractal differentiation, or fractal cycle is operating within each time segment. In examining the evolution of LIS, we pay attention to the developments within each discipline and its affiliated professional organizations to determine the effects on information science education. Using document analysis and interviews, we scrutinize the formation, dissolution and reformation of these professional organizations in response to the corresponding technological and social challenges faced in each segment.

In addition to examining professional organizations, we will examine the emergence of the iSchool movement, and determine its convergence and divergence from LIS. We will investigate the formation of the iField in response to the corresponding technological and social challenges faced in each segment. Aspects of examination will include changes in school names, repositioning of schools' missions and visions, changes in the need for faculty expertise, and evolution of ideas through research method and scope.Page Numbering, Headers and Footers

Do not include headers, footers or page numbers in your submission. These will be added when the publications are assembled.

4. REFERENCES

[1] Machlup, F., and Mansfield, U., and Peterson, Eds. 1983. *The study of information: Interdisciplinary messages*. John Wiley & Sons.

[2] Abbott, A. 2001. *Chaos of disciplines.* University of Chicago Press.

Building Folk UMLS: An Approach to Finding Meaning of Folk Terms in Medical Domain

Miao Chen
School of Information Studies, Syracuse University
mchen14@syr.edu

Bei Yu
School of Information Studies, Syracuse University
byu@syr.edu

Xiaozhong Liu
School of Information Studies, Syracuse University
xliu12@syr.edu

ABSTRACT
As a medical domain knowledge base, the Unified Medical Language System (UMLS) focuses on formal and professional medical terms; online health forums contain user-generated "folk terms", which can be used to complement the UMLS vocabulary. In this paper, we propose an approach to detecting folk terms from online discussions and matching their meanings to UMLS concepts. This approach makes connections between expert-built ontology and user-generated taxonomy (folksonomy) based on term distance matching using Google distance measurement. By finding meanings of user-generated folk terms, we will build what we call "folk UMLS" ontology as enrichment to the formal UMLS ontology.

Keywords
Ontology, concept extraction, ontology concept matching

1. INTRODUCTION
The UMLS ontology is a rich organized collection of medical terms and their semantic relations, covering a broad range of knowledge in the medical domain. On the one hand, the ontology contains a great number of vocabularies contributed by domain experts in a structured format; on the other hand, in the process of information use and interaction users generate folk terms, which are in unstructured format. Both the expert terms and the folk terms reflect knowledge of medical domain, while from different perspectives. Take folk phrase "later stage breast cancer" for example, it appears in user discussion board but is not listed as a concept in UMLS. In UMLS breast cancer stages is described by phrases like "stage I breast cancer" etc, which is a more rigorous way of categorizing stages. If we can combine the folk terms and the expert terms, we can build a more comprehensive knowledge base. Motivated by this goal, in this paper we introduce a new approach to building "folk UMLS", by connecting UMLS ontology and folk terms.

In this approach we need to solve two main problems: 1) identify and extract folk terms from user generated text; 2) map folk terms to the corresponding expert terms in UMLS. There have been previous trials on extracting concepts from textual data, with many of them statistically and machine learning based. Lin & Pantel (2002) presented a concept discovery method by automatically clustering words based on semantic similarities. In Maedche & Staab (2000)'s study, words in a corpus with high tf-idf values were recognized as candidate concepts. Besides statistical method, lexicon-based, rule-based, and combined methods have also been applied for concept extraction (Cohen & Hersh, 2005), Krauthammer et al. (2000) and Gaizauskas et al. (2003).

After candidate concepts are extracted from text, they need to be matched to ontologies, either to light-weighted ontologies like taxonomy or comprehensive ones like UMLS. Researchers have proposed various kinds of approaches to matching concepts to existing ontologies. Zou et al. (2003) developed an algorithm to detect UMLS concept through permuting input text and applying syntactic and semantic filters. The MetaMap project matched noun phrases to UMLS concepts based on parsing result of free text, by computing scores of distances between UMLS concepts and original text (Aronson, 2001).

Our study takes a different approach, which considers term semantic relatedness in Google database (Cilibrasi & Vitanyi, 2007), to linking ontology to folk terms using the Google distance measure. In the following sections, we will discuss how to use Google distance in our approach in details, and then address experiments and evaluation issues, and at last summarize the paper.

2. METHOD
Our approach consists of two phases. In the first phase, we take three steps to extract folk terms from online health forums:

1) Use natural language processing tool to chunk text to phrases;
2) Extract noun phrases from the phrase chunks, select the ones with high frequencies;
3) Search these noun phrases in UMLS; if a noun phrase is not an item in UMLS and is not a stop word, then it is considered as a candidate folk term.

The performance of term extraction is evaluated after the first phase. Non-expert users will be asked to judge whether the terms are medical related or not. We then proceed to the next phase to find meanings of these folk terms. The semantic distance between a folk term and a UMLS concept is defined by their Google distance. After computing its distance to all UMLS concepts, a folk term is matched to its closest UMLS concept.

The normalized semantic relatedness between two entities is computed using the Google distance in Cilibrasi and Vitanyi (2007)'s research. It is a co-occurrence based measure of term similarity. More specifically, given one webpage containing one term, Google distance measures the probability that it contains the other term (Gligorov et al., 2007). In the "later stage breast cancer" example, this is a folk noun phrase in medical forum and our goal is to project this entity to the UMLS concepts.

It costs high to compute the semantic distance between a folk term to all UMLS concepts. We take the following approach to reduce the computing load.

A folk term could be matched to a UMLS concept in two possible ways. The first one is easier by sharing the same head noun, like "flu" in both "pig flu" and "swine flu". The other one is more difficult, that is, they do not share the head noun but are semantically equivalent, e.g. "blood sugar" and "Glucose". In this study, we focus on the first matching type and leave the second matching type for future work. Below we use the "later stage breast cancer" as example to describe steps of the matching process:

1) Identify the head noun in the folk term "later stage breast cancer". We could compute the probability of each word in UMLS, like *P("cancer") > P("stage")*. Starting from the word "cancer" we find the largest possible match in UMLS. Here the largest phrase is "stage breast cancer".
2) Then find all concepts from UMLS containing the phrase as candidate matched concepts, i.e. "stage I breast cancer", "stage II breast cancer", "stage IV breast cancer", etc.
3) Compute the Google distance between the folk term and each UMLS candidate concept by using:

$$NGD(phrase, concept) = \frac{\max\{\log f(phrase), \log f(concept)\} - \log f(phrase, concept)}{\log M - \min\{\log f(phrase), \log f(concept)\}}$$

Where f(x) is the Google returned number of hits of phrase x and M is the number of total indexed pages by Google. If the target phrase and concept never occur together on the same web page, but do occur separately, the normalized Google distance (NGD) between them is infinite. If both the phrase and concept always occur together, then their NGD is zero.
4) Select the UMLS concept that has the lowest similarity with the phrase and match them. In this example folk phrase "later stage breast cancer" is matched to UMLS concept "stage IV breast cancer".

3. EXPERIMENT & EVALUATION

In our experiment, the data set comes from health forums such as WebMD, healthboards.com, and breastcancer.org, where people discuss medical related issues. We assume that people are more likely to use folk terms in such informal settings.

We will use the OpenNLP package, which is an open source natural language processing toolkit for finding noun phrases. It will detect sentence boundaries in free text and chunk sentences to phrases. The evaluation of folk term extraction will apply mechanism of precision and recall rates, as are frequently used in automatic term recognition. The precision is the rate of correctly identified terms among all identified terms, and the recall rate is the rate of correct terms in a document (Krauthammer & Nenadic, 2004). Similarly, the performance of concept mapping will also be measured by precision and recall. The first evaluation will obtain judgment from non-experts and for the second one we will ask medical experts to decide the matching performance.

4. SUMMARY

The paper proposes a new approach to discovering folk terms and connecting folk terms and professional ontologies by using Google distance measurement. By linking them, meanings of folk terms are made explicit thus can be further used in text processing tasks. It can facilitate medical document indexing by providing more related folk terms; it can be processed to Linked Open Data format to connect to other knowledge bases like UMLS; it can also provide a matching table to doctors who and patients and may enhance their communication quality. In the future, we will explore more possibilities to use the derived mapping between folk terms and formal terms as well as evaluate their usage.

5. REFERENCES

[1] Aronson, A.R. (2001). Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proceedings of AMIA Symposium 2001, 17-21.

[2] Cilibrasi, R. L., and Vitanyi, P. M. B. (2007). The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering, 19(3), 370-383.

[3] Cohen, A.M., and Hersh, W.R. (2005). A survey of current work in biomedical text mining. Briefings in Bioinformatics, 6(1), 57-71.

[4] Gaizauskas, R., Demetriou, G., Artymiuk, P.J., & Willett, P. (2003). Protein structures and information extraction from biological texts: the PASTA system. Bioinformatics, 19(1), 135-143.

[5] Gligorov, R., Aleksovski, Z., ten Kate, W., & van Harmelen, F. (2007). Using Google distance to weight approximate ontology matches. The 17th World Wide Web Conference 2007.

[6] Krauthammer, M., and Nenadic, G. (2004). Term identification in the biomedical literature. Journal of Biomedical Informatics, 37(6), 512-526.

[7] Krauthammer, M., Rzhetsky, A., Morozov, P., & Friedman, C. (2000). Using BLAST for identifying gene and protein names in journal articles. Gene, 259(1-2), 245-252.

[8] Lin, D., and Pantel, P. Concepts discovery from text .Proceedings of the 19th International Conference on Computational Linguistics, 577-583.

[9] Maedche, A., and Staab, S. (2000). Mining ontologies from text. Proceedings of the 12t European Workshop on Knowledge Acquisition, Modeling and Management, 189-202.

[10] Unified Medical Language System (UMLS). http://www.nlm.nih.gov/research/umls/

[11] Zou, Q., Chu, W.W., Morioka, C., Leazer, G.H., & Kangarloo, H. (2003). IndexFinder: A method of extracting key concepts from clinical texts for indexing. AMIA Annual Symposium Proceedings, 763-767.

Enhancing Access to the Web: Vocabulary Analysis on Users' Tags and Professionals' Index Terms

Yunseon Choi
Graduate School of Library and Information Science
University of Illinois at Urbana Champaign
501 E. Daniel Street, MC-493
Champaign, IL 61820-6211, USA
ychoi10@illinois.edu

ABSTRACT

This ongoing research aims to answer whether user-generated tags through social tagging could be used to enhance access to web resources and provide additional access points beyond professionally-generated ones. This study conducts qualitative vocabulary analysis of both users' tags and professionals' index terms.

Categories and Subject Descriptors
H.3. **[Information Storage and Retrieval]:** Content Analysis and Indexing – *indexing methods, linguistic processing, thesauruses*

General Terms
Performance, Human Factors, Standardization, Languages, Verification

Keywords
Controlled vocabulary, Digital Libraries, Folksonomy, Organization, Subject gateways, Subject indexing, Social tagging, Tags, Vocabulary analysis, Web

1. INTRODUCTION

A growing number of web resources have required new tools for organizing and providing access to the web. Subject gateways are such tools, designed to provide access to quality resources selected and indexed by specialists. However, a problem with these approaches is that most of them use traditional library schemes based on controlled vocabulary for subject access. Controlled vocabularies impede continuous development due to the rapid growth of digital libraries, so traditional indexing methods face the challenge in dealing with web resources. Furthermore, current systems are organized and indexed by professional indexers. Despite efforts to involve users in developing organization systems, these systems are not necessarily based on users' real languages.

iConference 2010, February 3–6, 2010, University of Illinois at Urbana-Champaign, Illinois, U.S.A.

Social tagging has received significant attention since it helps organize contents by user-generated tags. Social tagging allows users to add their tags to reflect their interests. Several researchers have discussed social tagging behavior and its usefulness for classification or retrieval. Nevertheless, further research is needed to qualitatively investigate social tagging and to verify its efficacy and benefit.

This paper is part of an ongoing research study which aims to answer whether and how social tagging could enhance access to web resources. In this paper, we provide the preliminary analysis of the following points: (1) whether tags have attributes beyond describing subjects of a document, (2) whether professional indexers have various or alternative interpretations of the same web document, and (3) whether tags would provide additional access points beyond index terms or keywords.

2. BACKGROUND

2.1 Organization of the Web

2.1.1 Subject gateways as organizing tools for the web

Subject gateways can range from "loosely collated commercial directories" such as Yahoo! subject categories, to "collections of quality assessed web resources compiled by the academic or research community" [1]. This study will refer to the concept of the latter for further discussion. Examples of such subject gateways include BUBL [2] and Intute [3]. BUBL describes itself as 'Free User-Friendly Access to selected internet resources covering all subject areas, with a special focus on Library and Information Science' [4]. It offers broad categorization of subjects based on the Dewey Decimal Classification (DDC) scheme [2]. Intute is a free web service aimed at students, teachers, and researchers in UK further education and higher education [3]. It is reported that Intute mainly uses the Universal Decimal Classification (UDC) and DDC for classification and has adapted them for in-house use. Intute also uses several thesauri for its subject relevance and comprehensiveness [5].

2.1.2 Challenges of controlled vocabulary for the web

For effective indexing, the indexing process needs to be controlled by using a so-called controlled vocabulary [6]. Yet, as there are more and more resources available on the web, existing controlled vocabularies have been challenged in their ability to index the range of digital web resources, e.g., slowness of revision, expensive indexing, and terms limited to topics found in physical and traditional library collections.

2.2 Social Tagging

Social tagging is described as "user-generated keywords" [7]. Since tags indicate users' perspectives in indexing resources, they have been suggested as a means to improve search and retrieval of resources on the web. Social tagging is a promising way to compensate for the disadvantages of traditional professional indexing because it is low-cost with a great number of users from everywhere contributing to the creation of tags. Thus, users' tags might be alternative terms for additional entry points of retrieval which are not easily attained using controlled vocabularies [8][9][10].

3. DATA COLLECTION

In order to examine professional indexers' vocabulary and compare it with that of users', we investigate two major subject gateways: BUBL and Intute (see Table 1). Both cover various subjects, and this feature allows one-to-one comparison on each subject area. We also extract tags from a social bookmarking site, Delicious.com. Unlike other social bookmarking sites, which provide the number of votes or users' comments, Delicious.com provides tagging data since it allows users to add, organize and share tags. Additionally, Delicious.com consists of a broad range of web resources, not limited to scholarly documents (e.g., journal articles on CiteUlike.org) or specific types of resources (e.g., photos and videos on flickr.com).

Table 1. BUBL vs. Intute

Site characteristics	BUBL	Intute
Classification	DDC	UDC and DDC
Keywords	N/A	*Controlled*: Several thesauri for their subject relevance and comprehensiveness, e.g., SCIE for Social Welfare, the Hasset, IBSS, LIR for Law, and the NLM MeSH headings for Medicine *Uncontrolled*: terms from web sites' titles and descriptions the indexers provide
Subjects covered	Various subjects	Various subjects
Database	Searchable and browsable	Searchable and browsable

Sampling documents is based on 10 subject categories BUBL provides as top-level categories (see Table 2). Under each category, documents in alphabetical order will be searched in turn at the other two sites, Intute and Delicious.com. Tags that are assigned to the document at Delicious.com are extracted only if a web document is found at all three locations (BUBL, Intute, and Delicious.com). Furthermore, indexers' index terms of both BUBL and Intute are collected for the comparison with users' tags.

Table 2. BUBL subject categories

Top Categories	Subjects covered
000 Generalities	Computing, Internet, Libraries, Information Science
100 Philosophy and psychology	Ethics, Paranormal phenomena
200 Religion	Bibles, Religions of the world
300 Social sciences	Sociology, Politics, Economics, Law, Education
400 Language	Linguistics, Language learning, Specific languages
500 Science and mathematics	Physics, Chemistry, Earth Sciences, Biology, Zoology
600 Technology	Medicine, Engineering, Agriculture, Management
700 The arts	Art, Planning, Architecture, Music, Sport
800 Literature and rhetoric	Literature of Specific languages
900 Geography and history	Travel, Genealogy, Archaeology

4. PRELIMINARY DATA ANALYSIS

One example of the analysis to be undertaken for each web resource in the sample is provided in this section. The poster will present findings from several more cases. Vocabulary analysis is conducted on the following main points: (1) analysis on Delicious.com tags, (2) analysis on BUBL and Intute vocabularies, and (3) analysis on Delicious.com tags and Intute keywords

(1) Analysis on Delicious.com tags

The process of identifying bibliographic attributes of tags is based on the Functional Requirements for Bibliographic Records (FRBR) model. Since the attributes defined in the FRBR model were derived from "a logical analysis of the data that are typically reflected in bibliographic records" [11], it would support a more systematic and meticulous analysis on the attributes of tags. A preliminary analysis of pilot data has identified that tags have several types of attributes beyond describing subjects of documents. The identified tag attributes can be categorized by the attributes defined by FRBR as shown in Table 3.

Table 3. Identified tags and related FRBR attributes

Identified tags	FRBR attributes
References or resources, research paper (tagged as "researchpapers"), article, tutorial, magazine, books or e-books, journal etc.	Form of work
Kids, children, senior, older, K-12 etc.	Intended audience (Work)
Audio, images, text etc.	Form of expression

(2) Analysis on BUBL and Intute vocabularies

In order to examine different points of view on the same document between professional indexers, indexers' index terms from BUBL and Intute are analyzed. BUBL offers each document with the classification number based on DDC. For indexer's index terms from BUBL, this study analyzes index strings, which are category paths of classification. For example, regarding a document, *Amazon.com*, the following category paths can be recognized, and they will be collected for analysis:

- News media, journalism, publishing > Publishers and publishing > Booksellers and bookshops

The collection of an indexer's index terms from Intute is the same as BUBL. For a more accurate comparison based on an equal condition, only index strings of category paths in classification schemes are analyzed:

- Communication and Media Studies > New Media > Interactive Games and Gaming
- Music > Music Industry, Recording and Publishing
- Communication and Media Studies > Publishing > Bookselling

(3) Analysis on Delicious.com tags and Intute keywords

In order to inspect whether Delicious.com users' tags would provide additional access points beyond index terms or keywords that Intute professional indexers provide, the top ranked tags assigned to a document at Delicious.com are collected and normalized. This is done through the rules for vocabulary analysis such as checking spelling and word forms. The top 10 tags are compared with keywords (controlled or uncontrolled) from Intute. Intute's uncontrolled keywords are added if its indexers can find no suitable word in thesauri. The keywords provided by Intute are useful and are the most appropriate data in order to compare the professional indexer's point of view with the user's point of view in subject indexing on the same document.

Table 4. Intute Keywords vs. Delicious Top 10 tags

Keywords at Intute		**Tags at Delicious.com**
Keywords - controlled	Amazon.com (Firm); books; publishing; publishers; bookselling; booksellers; electronic publishing; bookstores; motion pictures (visual works); videotapes; video games; digital versatile discs; music; software	shopping, books, amazon, online, bookstore, music, web, internet, fun, deals
Keywords - uncontrolled	online; electronic commerce; on-line; book stores; bookshops; e-publishing; films; movies; motion pictures; video tapes; digital video discs; DVDs; compact discs; CDs	

5. DISCUSSION

Table 3 illustrates that tags provide essential bibliographic attributes, which have not been identified in previous research. This provides a helpful understanding of features and patterns of tags in describing web documents.

Moreover, the preliminary analysis has revealed that there were some various or alternative interpretations in new subject areas, for instance, internet-related areas. There were different perspectives on the same document, *Amazon.com*, even between groups of professional indexers. BUBL places it at the category of *070.5 Publishers and Publishing* under the category of *070 News media, Journalism, Publishing*. Intute classifies it as the similar subjects with BUBL, e.g., *New media* or *Publishing*. However, Intute also categorizes it at the category of *Music industry, recording and publishing* under the category of *Creative and performing arts.*

Table 4 indicates that among the top 10 tags at Delicious.com, a term "shopping" which is ranked first is not included in the Intute keywords. However, it is worthwhile to note that the tag "shopping" might be an additional helpful access point for those who are interested in purchasing books or other related goods online.

6. CONCLUDING REMARKS

As part of an ongoing research study, this paper focuses on bridging the gap of insufficiency of studies on vocabulary analysis by comparing user-generated tags with professional-generated index terms regarding web resources. Current work will be complemented by quantitative measures performed on a large data set. The research also will evaluate indexing consistency of tagging and professional indexing in order to systematically verify the efficacy and quality of tags. This will provide a way of improving the organization of web resources by increasing the utilization of social tagging data.

7. ACKNOWLEDGEMENTS

This research has been conducted under the direction of Professor Linda C. Smith at the University of Illinois at Urbana-Champaign. I wish to express my deepest respect and gratitude to her.

8. REFERENCES

[1] University of Kent. *Library Services Subject Guides*. Retrieved from http://www.kent.ac.uk/library/subjects/healthinfo/subjgate.html. 2009

[2] BUBL Link Home. http://www.bubl.ac.uk

[3] "Intute." Wikipedia, The Free Encyclopedia. 2009. Wikimedia Foundation, Inc. 10 Sep. 2009.

[4] "BUBL." Wikipedia, The Free Encyclopedia. 2009. Wikimedia Foundation, Inc. 10 Sep. 2009.

[5] *Personal Communication via email with Intute.* May 21and June 2, 2009.

[6] Lancaster, F. W. 1972. *Vocabulary control for information retrieval*. Washington, D.C.: Information Resources Press.

[7] Trant, J. 2009. Studying Social Tagging and Folksonomy: A Review and Framework. *Journal of Digital Information* 10(1).

[8] Maltby, A. 1975. *Sayers' Manual of Classification for Librarians* (5th Ed.), Andre Deutsch, London.

[9] Hayman, S. 2007. Folksonomies and tagging: New developments in social bookmarking. *Ark Group Conference:*

Developing and Improving Classification Schemes 27-29 June, Rydges World Square, Sydney
[10] Quintarelli, E. 2005. "Folksonomies: power to the people", *Proceedings of the 1st International Society for Knowledge Organization* . UniMIB Meeting, June 24, Milan, Italy, ISKOI, Italy.
[11] IFLA Study Group. 1998. *Functional requirements for bibliographic records: final report.* München: K.G. Saur.

Leaders Wanted: Mentoring and Retaining Librarians of Color

Nicole A. Cooke, MLS, M. Ed.
School of Communication & Information
Rutgers University
4 Huntington St., New Brunswick, NJ 08901
Nicole.Cooke@Rutgers.edu

ABSTRACT

In order to successfully mentor, and hopefully retain, new professionals in the field, especially those from underrepresented populations, we need to know what professional information they are seeking and deem valuable. Using the 2009 American Library Association (ALA) Spectrum Scholar Leadership / Reach 21 Institute as a specific example, this poster explores the information needs of current MLIS students and brand new MLIS graduates, and provides insight into what this group wants and needs to know from experienced professionals.

In July 2009, 50 ALA Spectrum Scholars (masters level scholarship recipients) and 20 Reach 21 Scholars (students from other minority library science scholarship initiative) attended the institute as a supplement to their academic scholarships. The IMLS-funded project, ***REACH 21: Preparing the Next Generation of Librarians for Leadership***, builds on Spectrum's past accomplishments and extends the community and support benefits of the program to even greater numbers of future librarians. The leadership institute and REACH 21 aim to foster the recruitment, matriculation, and early career development of racially and ethnically diverse students in master's-level library and information studies programs, provide mentoring, coaching and support networks for these students, and aid in educational and early career retention. Institute participants represented a variety of iSchool member institutions including UCLA, Rutgers University, Drexel University, Florida State University, University of North Texas, University of Illinois, Indiana University, University of Maryland, University of Pittsburgh and University of Washington.

In 1998 social learning theorist Etienne Wenger wrote the book *Communities of Practice*, in which he asserts that learning is a social process and is based upon participation; it is this participation that allows us to form and engage in communities. "Such participation shapes not only what we do, but also who we are and how we interpret what we do" [6]. Wenger describes communities of practice (COPs) as being everywhere and encompassing all people, students, family members, coworkers, professionals, etc.

Mentoring, whether formal or informal, is an important outcome of professional COPs; there is a large body of literature on mentoring that can be applicable to professional communities of practice. The literature suggests that mentoring is a reciprocal growth process, benefiting the mentor and the mentee, and is a good vehicle for developing leadership skills. "Growth, even highly desired and positive, is not easy. Leadership growth is a process where personal paradigms are challenged and pushed beyond one's comfort zone. Mentoring helps soften the discomfort and provides caring and helpful individuals for encouragement and support" [4]. Mavrinac (2005) concurs by stating that mentoring is an inclusive, democratic and motivating relationship that can "serve to widen an employee's learning context within and outside the organization" [5].

Using Brenda Dervin's concept of sense-making to frame this research project, the goal is to identify the information needs of new MLIS graduates so that designated mentors can provide appropriate and valuable information to help them succeed in their new professional positions. It is theorized that new library professionals can benefit from the information and expertise of those already in the field; once students graduate they no longer receive parceled and structured information from their professors, rather they must take charge of their own continuous learning, and begin to "learn on the job." At this juncture, students / new professionals can encounter information gaps. Dervin refers to these gaps as knowledge gaps, information inequities, communication gaps, information voids and information deficits [2]; these gaps represent points at which discontinuities occur, or points at which information ceases to flow and the individual cannot move forward without "constructing a new or changed sense" [1]. Individuals must construct new meanings, based on their contexts and situationality, in order to learn and continue intellectual expansion.

Sense-making provides a frame with which to study the information needs, seeking and use of this group [3], and provides a "set of methods which have been developed to study the making of sense that people do in their everyday experiences" [1]. Specifically, sense-making guides the researcher to ask when the information gap occurs, what the nature of the information gap is, and what information will suffice the gap. In this particular research project, the gap occurs when new library professionals straddle the time period between graduation and entry into a new job. The nature of the gap occurs during a professional and academic leadership workshop in which the MLIS students / new graduates were asked what future workshops and seminars they would find valuable, and their own responses reveal the type of information required to fulfill their information gaps.

Over the course of the three day leadership institute, the students were presented with a variety of workshops and seminars covering a wide range of topics applicable to library leaders, including interviewing skills, resume development, community involvement, diversifying the library and finding a mentor. At the conclusion of the institute, participants completed a survey which asked, in part, "If we could have incorporated more time for any one activity or topic, what would it have been?" In conjunction with the responses provided in the "additional comments" section, this qualitative data was analyzed to determine what programs should be developed for future institutes, and to ascertain what information the group deems necessary and valuable as they prepare to enter their first library positions.

In order to determine the information necessary to suffice this information gap, the constant comparison method was employed to analyze the qualitative data provided by the surveys. The data were compared and contrasted, and as a result 3 categories of responses emerged and were defined: (1) topics the participants would like to explore further, (2) activities the participants enjoyed and found useful, and (3) outcomes (lessons and goals achieved during the institute). The analysis indicated that information related to information about the profession (specialties, professional organizations), outreach and advocacy, job skills (navigating the job search, resume and cover letter preparation, and interviewing), networking, and increasing diversity in the profession, are the information corpora that will suffice the information gaps of graduate students in library and information studies and new librarians.

With this information in mind, those who choose to serve as mentors to this demographic of librarians will be armed with appropriate information to best serve their mentees, hopefully providing them with the knowledge and support that will encourage their retention in the field. Among the goals of this research is to cater to a specific group (ALA Spectrum and Reach 21 Scholars), while gleaning insight and knowledge that will benefit larger populations of new library graduates, and inform the literature on mentoring.

REFERENCES

[1] Dervin, B. From the mind's eye of the user: the sense-making qualitative-quantitative methodology. In Glazier, J. D. and Powell, R. R. eds. Qualitative research in information management. Libraries Unlimited, Englewood, CO, 1992, 61-84.

[2] Dervin, B. Communication gaps and inequities: Moving toward a reconceptualization. In Dervin, B. and Voight, M. J. eds. Progress in communication sciences. Ablex, Norwood, NJ, 1980, 73-112.

[3] Dervin, B. and Nilan, M. Information needs and uses. Annual Review of Information Science and Technology, 21(1986), 3-33.

[4] Golian-Lui, L. M. Fostering librarian leadership through mentoring. Adult Learning, 14, 1 (2003), 26-28.

[5] Mavrinac, M. A. Transformational leadership: Peer mentoring as a values-based learning process. portal: Libraries and the Academy, 5, 3 (2005), 391-404.

[6] Wenger, E. C. Communities of practice: Learning, meaning, and identity. Cambridge University Press, NY, 1998.

Digital Preservation Education in iSchools

Kaitlin L. Costello
School of Information & Library Science
University of North Carolina
Chapel Hill, NC USA
+1 919-627-1741
kaitcost@email.unc.edu

ABSTRACT

This poster investigates digital preservation education in the iSchool caucus. The project identifies core concepts addressed in digital preservation coursework in iSchools and identifies possible areas for curriculum development.

Digital preservation education at the graduate level is critical. To ensure long-term access and use of digital materials, information professionals must have a working knowledge of digital curation, which emphasizes a lifecycle approach to digital preservation [1]. Unfortunately, the topic of digital preservation education is not prominent in literature about digital curation. Only a handful of case studies and recommendations have been published regarding digital preservation education within information science, library science, and computer science graduate programs. Instead, much of the work on digital preservation education is contained in more general studies on educating digital librarians or electronic records managers. To understand how to better design curricula that engages central issues of digital curation at the graduate level, an investigation of the current state of digital preservation education is warranted.

Coursework devoted solely to digital preservation is essential for graduate students in information-centric disciplines. The necessity for devoted coursework is due to the complex and multifaceted nature of the topic. Unfortunately, a 2006 study found that very few library or information science schools offered courses specifically on the topic of digital preservation. Furthermore, an extremely small percentage of students in library or information science programs had exposure to the critical aspects of digital preservation during their coursework [2].

Digital preservation education can and should be studied in iSchools. The core mission of the iSchool movement is to connect people, information, and technology [3]. Digital curation supports this mission by enabling the continued maintenance of digital information resources throughout their lifecycle, allowing them to be rendered and re-used in the long-term. It is an interdisciplinary process that hinges on expertise from many different fields, including computer science, information and library science, informatics, management, and education. Furthermore, iSchools are a natural home for digital library education [4] and there are significant overlaps between digital library education and digital curation education [5]. It follows that iSchools are an excellent venue for research on the topic of digital preservation education.

This project examines digital preservation courses in iSchools over the past five years (2005-2009). Course descriptions and syllabi are examined in order to develop a definition of current practices in digital preservation education. Based on this definition, areas for future developments in digital preservation curricula are identified.

Course catalogs from the 26 iSchools have been analyzed to determine whether or not schools offer classes specifically on the topic of digital preservation. Of the 26 iSchools, 9 schools offer degrees in information science and in library science, 6 award degrees in information science but not in library science, and 5 award degrees in library science and not information science. The remaining 6 schools offer a variety of degrees, including computer science, information management, and information technology. These categories will be useful in determining what types of iSchools, if any, are leaders in digital preservation education. All of the schools that have been examined to date offer course catalogs and course descriptions on the open web. Many of the course syllabi are also available online. The course must contain the phrase "Digital Preservation" in its title or course description in order to be included. One-shot sessions and classes that deal with a subset of digital preservation, such as classes on digital libraries, are not considered.

Course themes and assignments are compared to the DigCCurr Matrix of Digital Curation Knowledge and Competencies. This six-dimensional matrix from the University of North Carolina DigCCurr project defines and organizes materials to be covered in digital curation coursework [6]. This analysis will identify current strengths and potential areas for further development in digital preservation education. The study will also address the question of where current digital preservation course materials fit within the larger scope of digital curation knowledge and competencies.

Categories and Subject Descriptors

K.3.2 [**Computers and Education**]: Computer and Information Science Education – *curriculum.*

Keywords

Digital curation, digital preservation, graduate education, iSchool

1. ACKNOWLEDGMENTS

This work was supported through IMLS Grant #RE-05-08-0060-08. I am grateful to my anonymous reviewers, Dr. Cal Lee, Dr. Helen Tibbo, Sarah Ramdeen, and Mike Brown for their valuable suggestions.

2. REFERENCES

[1] Joint Information Systems Committee. (2003). An invitation for expressions of interest to establish a new Digital Curation Centre for research into and support of the curation and preservation of digital data and publications. Retrieved

October 20, 2009. http://www.dcc.ac.uk/docs/6-03Circular.pdf.

[2] Gracy, K. R., & Croft, J. A. (2006). Quo vadis, preservation education? A study of current trends and future needs in graduate programs. *Library Resources & Technical Services 50*(4), 274-94.

[3] iSchools Caucus. (2009). About the iSchools. Retrieved October 20, 2009. http://www.ischools.org/site/about/.

[4] Wildemuth, B., Pomerantz, J., Oh, S., Yang, S., & Fox, E. (2009). I-schools as a natural home for digital libraries education. Presented at the iConference 2009, Chapel Hill, NC. http://nora.lis.uiuc.edu/images/iConferences/DL_in_i-schoolspaper-final2009-01-27.pdf.

[5] Pomerantz, J., Oh, S., Wildemuth, B., Hank, C., Tibbo, H., Fox, E., et al. (2009). Comparing curricula for digital library and digital curation education. In *Proceedings of DigCCurr 2009. Digital Curation: Practice, Promise & Prospects*, Chapel Hill, NC, April 2009. H. Tibbo, C. Hank, C. Lee, and R. Clemens, Eds. SILS: Chapel Hill, NC, 2-3.

[6] Lee, C. (2009, June 17). Matrix of digital curation knowledge and competencies, Version 13. Retrieved October 26, 2009. http://www.ils.unc.edu/digccurr/digccurr-matrix.html.

Work in Progress: What is "Enough"?

Jeanette de Richemond
School of Communication and Information
Rutgers, The State University
4 Huntington Street
New Brunswick, NJ 08901-1071
0011 - 1 – 215-262-3725

jderichemond@gmail.com

ABSTRACT

This poster presents dissertation work in progress on the question of "enough." The research focus is the assessment of "enough" information to make a decision, in particular a medical decision determining the diagnosis of a patient. "Enough" is considered "enough" information to facilitate making a decision or taking an action. Qualities of qualities of "enough" are identified and described by analyzing case reports published in the *New England Journal of Medicine*. Findings are reported, and contribute to the development of a conceptual model of factors contributing to "enough.'

Categories and Subject Descriptors

D H.1.1 (**Systems and Information Theory)**

General Terms

Human Factors

Topics

Information Seeking and Use, Health Informatics

Keywords

Enough, Information Behavior, Clinical Informatics

1. Introduction

The focus of dissertation research is the assessment of "enough" information to make a decision, in particular a medical decision determining the diagnosis of a patient. In information science, the determination of "enough" information to traverse gaps (Dervin, 1992) and make progress has important implications for the design of information retrieval systems, particularly the presentation of retrieved results. "If the United States is to realize the full value of biomedical knowledge..., the mechanisms through which that knowledge is operationalized and care is delivered must be radically redesigned." (Shekelle, Morton, & Keeler, 2006, p. 28). Assessing "enough" information to make a decision is intrinsic to efficient and effective use of clinical information.

1. Objective

The objective of my research is to examine and to explore the assessment of "enough." For the purposes of this project, "enough" is considered enough information to facilitate making a decision or taking an action by an individual or a team. The goals of the initial phase of qualitative research, which involved studying standardized case reports published in the New England Journal of Medicine (NEJM), are to:

1.1 Describe specific characteristics of "enough" as revealed in the process of making a diagnosis.

1.2 Identify problematic situations or patient cases, which have similarities, and in which similar qualities of "enough" resolve the situation.

1.3 Identify problematic situations that vary, and in which "enough" presents unexpected characteristics to resolve the situation.

1.4 Develop a conceptual model for describing the inter-relationships of the characteristics of the problematic situation and related work tasks and information behavior that influence the medical decision-making.

2. Methods

The medical arena provides an excellent ground for the study of "enough" as the specific actions taken to achieve "enough" are specifically documented in patient records, and "enough," the equivalent to a diagnosis in this study, is clearly delineated in patient records.

Case reports published in the New England Journal of Medicine provide the initial information for analysis. Each case report is broken down into episodes of care; each episode of care typically includes a variety of work-subtasks. Written descriptions of each case report are prepared, consisting of four elements: a description and discussion of the problematic situation, a description and classification of the work task and its sub-tasks, a description and discussion of the linked information behavior, and an interpretation of "enough" in each episode of care.

3. Preliminary Results

Findings concerning the inter-relationships between the problematic situation, work task, and information behavior leading to "enough" contribute to the development of a conceptual model. (See Figure 1.) The conceptual model demonstrates the interaction of factors involved in assessing "enough."

The model will be applied to characterizing and analyzing problems, work tasks, and information behavior in continued research on assessing "enough" information to make a medical diagnosis. (Note: The conceptual model will be included in the poster.)

Findings will also be used to extend the Li and Belkin (2008) task classification scheme to incorporate qualities of "enough."

(Note: Additional findings will be reported in the poster.)

4. Conclusions

The next phases of my dissertation research involve:

4.1 Developing a new methodology, which would determine how

to use clinical data to study an information science problem. This is a new approach to clinical informatics, involving use of clinical data repository as the information required to reach each decision can also be clearly defined in medicine. These elements, which are necessary in studying the assessment of enough, are recorded in a patient chart.

4.2 Conducting qualitative research in which a group of working physicians in a specialty such as cardiology will be presented with a case study of a patient along with a set of materials (patient history, reports of diagnostic tests, articles from medical journals, copies of recent medical journals, etc.) related to this patient as well as Internet access. Study participants will be asked to classify and evaluate each item the portfolio, and to select and determine what will be "enough" information to develop a diagnosis. Physicians will also be interviewed to obtain more information on their process assessing what is enough information to diagnose the patient.

REFERENCES

[1] Dervin, B. 1999. On studying information seeking methodologically: the implications of connecting metatheory to method. Information Processing and Management 35, 727-750.

[2] Li, Y. & Belkin, N.J. 2008. A faceted approach to conceptualizing tasks in information seeking. Information Processing and Management, 44, 1822-1837.

[3] Shekelle, P.G., Morton, S.C., & Keeler, E.B. 2006. Costs and Benefits of Health Information Technology. Evidence Report/Technology Assessment No. 132. (Prepared by Southern California Evidence-Based Practice Center/RAND under Contract No 290-97-0001.) Rockville, MD: Agency for Healthcare Research and Quality (AHRQ).

Figure 1: Conceptual Model of "Enough"

DMCA Take-down Notices on Campus: A Case Study

Wyatt Ditzler
University of Wisconsin-Milwaukee, School of Information Studies
3210 North Maryland Ave, Room 510
Milwaukee, WI 53211
1-414-229-4707
wditzler@uwm.edu

Michael Zimmer
University of Wisconsin-Milwaukee, School of Information Studies
3210 North Maryland Ave, Room 510
Milwaukee, WI 53211
1-414-229-3627
zimmerm@uwm.edu

Tomas Lipinski
University of Wisconsin-Milwaukee, School of Information Studies
3210 North Maryland Ave, Room 510
Milwaukee, WI 53211
1-414-229-4908
tlipinski@uwm.edu

ABSTRACT
The purpose of this study is to investigate the effects the Digital Millennium Copyright Act (DMCA), Section 512(c), take-down notices have on a university campus. Specifically this study will examine: the policies and procedures one university employs to comply with the DMCA, the content of received DMCA take-down notices, whether the notices comply with the standards set forth in the DMCA for notification, and the effects DMCA take-down notices have on university students and faculty.

Categories and Subject Descriptors
N/A

General Terms
Legal Aspects

Keywords
Digital Millennium Copyright Act, DMCA, take-down notices, university campus

1. INTRODUCTION
In 1998 President Bill Clinton signed into law the Digital Millennium Copyright Act (DMCA). Title II of the DMCA created Section 512, the 'Safe harbors' provision. Title II of the DMCA was a compromise between online service providers (OSPs) and copyright holders first negotiated in the Online Copyright Infringement Liability Limitation Act of 1998 [1]. Concerns over the 'Safe harbors' provision are continually highlighted by Chilling Effects, a partnership between groups such as: the Berkman Center for Internet & Society, DePaul University College of Law, and the Electronic Frontier Foundation to name a few [2]. Eleven years after codification, research detailing the affects that the 'Safe harbors' provision has created on university and college campuses is lacking. This case study proposes to answer the following questions: What policies and procedures does a university campus employ to comply with the 'Safe harbors' provision? Do the DMCA take-down notices, outlined in 512(c)(3), comply with the guidelines set forth in the law? What affects do DMCA take-down notices and university policies have on the education of university students? How do DMCA take-down notices affect the intellectual freedom of both students and faculty on campus?

2. METHODOLOGY

2.1 Case study site
The site of this case study is a state university located in an urban center located geographically in the Midwest of the United States. The university has a student enrollment around 30,000 and a faculty numbering around 1,500. The university provides on-campus housing for students or university owned housing scattered around the city. A high percentage of first year undergraduate students reside in university owned housing, however after the first year many students move to non-university owned housing. Each student is provided with network access on campus, website space, and file storage on university owned equipment accessible via the Web.

2.2 Access to DMCA take-down notices
Access to DMCA take-down notices may prove difficult for this research. The university may view the take-down notices as confidential information under the Family Educational Rights and Privacy Act or other university privacy policies when the subject of the notice is a student. Further, research concerning take-down notices at the university may raise concerns that the end result may negatively impact the reputation of the institution. Initial contact with university officials has been mixed. There are three university offices that deal with DMCA take-down notices, university IT, the legal office, and the Dean of students. Of the three university offices, only one has offered support and expressed interest in collaborating on the study, university IT. The other two university offices have not responded to an initial inquiry sent by the authors.

A second means to obtain access to the DMCA take-down notice is to submit an open records request. The state in which this university is located has a robust open records statute which would be applicable for this research. This method would provide the least favorable outcome for this research however. Important information from the DMCA take-down notices may be redacted

or worse, all but the basic language found in every DMCA take-down notice would be accessible. An outcome such as this would be severely limiting to the potential usefulness of the study.

2.3 Policy and process analysis

The authors will conduct a legal analysis of the university policies and procedures directly related to the receipt of a DMCA take-down notice. As part of the policy analysis, the authors will examine any programs the university employs to help students understand copyright law and their rights under copyright law. Especially of interest are any materials the university provides on the receipt of a DMCA take-down notice or how one may respond to a DMCA take-down notice. Later, interviews with appropriate members of the three university offices will be conducted. The interviews will focus on university policy and procedures related to their respective offices. Among the subjects of inquiry will be questions related to how the university developed policy, whether the DMCA take-down notice function is efficient or burdensome for the university, what ways the university is combating copyright infringement, and how to improve the 'Safe harbors' provision and take-down notices for the university. Finally, the authors will also conduct an ethical analysis on the university's policies and procedures.

2.4 DMCA take-down notice analysis

DMCA take-down notices will be coded in a restricted access database. Basic information related to the subjects of the DMCA take-down notice would be limited to classification of faculty or staff. However detailed information about the 'complaining party' would be kept for statistical inquiry. The DMCA take-down notices would also be checked against the 'Elements of Notification' found in 512(c) (3) for adherence to the prescribed notification process.

The DMCA take-down analysis will also include the identification of the allegedly infringing materials. As part of the notification process, the 'complaining party' was to identify the location of allegedly infringing materials. The authors will investigate what file types are most frequently targeted and try to reconstruct the context of the targeted file.

3. OUTCOME

The results of this research will begin to fill a void in the literature focused on the effects of the Digital Millennium Copyright Act. Policy decisions require a verbose amount of information in order to make well tailored and proper laws. In August of 2008, the Higher Education Opportunity Act was signed into law. This law, which will go into effect in July 2010, requires that universities and college campuses develop plans and employ technological deterrents to assuage, or ideally halt, copyright infringement on university and college campuses. Unfortunately this research was not made available in time for that policy deliberation. However any future policy deliberations may find the results of this research informative.

4. REFERENCES

[1] URBAN, JENNIFER M. AND LAURA QUILTER. 2006. Efficient process or "chilling effects"? Takedown notices under section 512 of the Digital Millennium Copyright Act. *Santa Clara Computer & High Tech. L.J. 621*.

[2] Chilling Effects http://www.chillingeffects.org/about

Participatory Media for Education: Driving Student-Centered Learning

Nathan Gandomi
UC Berkeley School of Information
University of California Berkeley
Berkeley, CA 94707
ngandomi@ischool.berkeley.edu

Erin Knight
UC Berkeley School of Information
University of California Berkeley
Berkeley, CA 94707
eknight@ischool.berkeley.edu

ABSTRACT

With the advent of Web 2.0 technologies, participation and collaboration have become predominant experiences on the Web. The teaching and learning community, as a whole, has been late to capitalize on these technologies in the classroom. How can we support pedagogical change with web-based course management systems and participatory media? Our research will attempt to answer this question through observation and analysis of faculty and student use of the tools in the course context, interviews with faculty and instructional support designers and a comprehensive research review. This poster reports some preliminary findings from a pilot study using the Social Media Classroom (SMC, 2009), a lightweight, open course site solution with embedded social media tools such as wikis, blogs, forums, chatrooms and social bookmarking, and will outline the following components: (1) Use of the Social Media Classroom in three I School courses, (2) Student/Instructor usage patterns, expectations, evaluations and best practices across courses, (3) Top-rated features we built to make the SMC more robust, which have since been released to the open community and (4) Implications for future research.

Categories and Subject Descriptors

K.3.1 [**Computers and Education**] : Computer Uses in Education - *Collaborative learning*

General Terms

Management, Measurement, Design, Human Factors, Standardization, Theory

Keywords

Teaching and learning, education technology, social media, participatory media

INTRODUCTION

In the last 15 years, Web technologies, such as Learning Management Systems (LMS), have shifted the traditional teaching and learning paradigm by extending classroom borders, capturing and persisting course content and giving teachers more flexibility and access to students and other resources. However, most of these technologies and systems also constrained and limited the evolution of teaching and learning by supporting a traditional, instructional framework. Each LMS simply enabled and guided teachers to provide ("upload") all the course materials. Students were still seen as the end-users or consumers of the information.

With the advent of Web 2.0 technologies, participation and collaboration have become predominant experiences on the Web. The teaching and learning community, as a whole, has been late to capitalize on these technologies in the classroom, perhaps because of uncertainty around how to incorporate them, or due to constraints imposed by the LMS. Additionally, part of the delay in uptake stems from the fact that participatory media tools require an additional shift in educational paradigms, from instructional, on-the-pulpit type of teaching, to a student-centered, adaptive environment where students can contribute to the course material and learn from one another.

Lately, there has been more and more buzz around the potential of these Web 2.0 tools and technologies to improve education. More people are exploring how the embedded ideas of user-generated content, network effects of mass participation, openness and low barriers to entry can be applied to traditional education axioms like student engagement, interaction in learning, and student ownership and management of learning. (Mason & Rennie, 2008) Additionally, there is increasing focus on the potential of student-centered learning paradigms and how these types of tools can support the shift. With roots in Papert's (1980) constructionism and Vygotsky's "Zone of proximal development" and apprenticeship models of learning (Rogoff, 1990), social and collaborative learning can enable students to construct a deeper understanding of material and lead to outcomes not possible in a strictly top-down learning environment.

How can we support pedagogical shifts using course systems and participatory media? Our research will attempt to answer this question through observation and analysis of faculty and student use of the tools in the course context, interviews with faculty and instructional support designers and a comprehensive research review. This poster reports some preliminary findings from a pilot study using the Social Media Classroom, a lightweight, open course site solution with embedded social media tools such as wikis, blogs, forums, chatrooms and social bookmarking, and will outline the following components:

- Use of the Social Media Classroom in three I School courses
- Student/teacher usage patterns, expectations, evaluations and best practices across courses
- Top-rated features we built to make the SMC more robust, which have since been released to the open community
- Implications for future research.

METHODS

We have implemented the Social Media Classroom (SMC) in three courses at the I School. The SMC is an open and freely accessible course site solution with embedded social media tools to support teaching and learning, and expand the course experience. The system was developed by Howard Rheingold in 2007 and formally released to the public in May 2009.

The SMC has two key features that position it as a powerful and potentially significant education solution for educators and researchers:

- **Openness** - open source, freely accessible and open educational content and resources
- **Embedded social media tools** - participatory tools including wikis, blogs, forums, chatrooms and social bookmarking, built directly into the course environment to empower social construction of knowledge and a student-centered learning environment

We have been observing usage patterns across all three courses, as well as a number of previous courses taught by Rheingold through the system. These courses are hybrid learning environments, where students attend a face-to-face class but also use the SMC to interact with the material and other classmates. The class sizes range from 30 to 40 students. We surveyed students at the beginning of the semester about their familiarities with the tools, as well as their perceptions around the educational value of each. We did a mid-semester survey to get evaluations of the site, as well as a more in-depth survey and interviews at the end of the semester to get a better understanding of the student experience with the system and individual tools.

We are also interviewing a number of faculty and teachers in the UC system to get a better understanding of obstacles and concerns around this type of technology, as well as developing a "readiness" measure to identify those instructors who are ready to incorporate a student-centered paradigm into their class, and how these tools and scaffolding materials around each can support them in doing so.

RESULTS

Our study is still in progress through the remainder of this semester. We will present detailed findings in February.

DISCUSSION / FUTURE DIRECTION

As previously mentioned, the teaching and learning community has been slow to adopt Web 2.0 participatory media tools into the classroom. We want to get a better understanding of why that is, and how to support those that are ready to adopt them. This requires a better understanding of multiple dimensions including, how the technologies are used by instructors and received by students, how to measure effectiveness of usage and best practices and scaffolding to support use. Based on our pilot study, interviews and research, we aim to further explore the following six areas, and provide our findings and future directions for each.

1. Patterns of use/Types of learners that emerge from these systems from collaborative learning environments
2. Types of learning/learning theory supported by this type of system (21st century skills, constructionist learning, etc.)
3. Student/Instructor expectations/evaluations
4. Effectiveness/Evaluation of participatory media tools for education
5. Supporting instructors in the shift to student-centered paradigm through these types of tools (scaffolding, best practices)
6. Model course site solution for hybrid and virtual learning (what tools to include, etc.)

REFERENCES AND CITATIONS

Hanson, P., & Robson, R. (2004). Evaluating course management technology: A pilot study. *Educause Center for Applied Research, Research Bulletin*, (24), Boulder, CO: EDUCAUSE. http://www.educause.edu/library/ERB0424

Kvavik, R., & Caruso, J. (2005). Study of students and information technology: Convenience, connection, control and learning (Vol. 6). Boulder, CO.: *Educause Center for Applied Research*, Research Study. http://www.educause.edu/apps/er/erm08/erm0740.asp

Mason, R., & Rennie, F. (2008). *E-learning and Social Networking Handbook*. New York: Routledge.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY: Basic Books.

Rogoff, B. (1990). *Apprenticeship in Thinking*. New York, NY: Oxford University Press.

[SMC] Social Media Classroom (2009). http://socialmediaclassroom.com

Exploring Methods in Community Informatics (poster)

Jeff Ginger, Adam Kehoe and Navadeep Khanal
University of Illinois at Urbana-Champaign
Graduate School of Library and Information Science
501 E. Daniel St.
Champaign, IL 61820
+12173333280
ginger@illinois.edu, kehoe@illinois.edu, khanal@illinois.edu

Topics

Information ethics, Research methods, Community informatics, Quantitative methods, Qualitative methods.

Keywords

Community Informatics, methods, cyberpower, community inquiry, participatory action research (PAR).

Poster Abstract

Community Informatics (CI) is an emerging field of study, practice and activism that has grown in popularity and influence in recent years. As an academic discipline CI is typically situated within iSchools and provides an important venue for their connection to community knowledge, educational practice, and social justice movements [1]. The term was originally brought into popular use by Loader and Gurstein in the late 90's and contrasted in relation to the overarching study of social informatics, which at the time was mostly concerned with business and government connections to information technologies [2]. As ICT's and cultures embedded in our information society have evolved, however, the lines between community, institutional, and individual ICT cultural practices have blurred; no longer can public computing be conceptualized as just a machine at the local library or can digital divide power inequities be cast as a simple lack of access to information. As result community informatics has become widely interpreted in terms of research, theory, methods and places of application.

This has given rise to a need for continued discussion over the definition and application of methods in community informatics. Our poster seeks to provide representation of some of the methodological perspectives encountered in a few projects of the Community Informatics Initiative (CII) [3], a research and teaching center and associated curriculum that is part of the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. Our work is far from conclusive, but instead intended to be a starting point for discussion about theories and examples of CI methods in action. Specifically, we address:

Participatory Action Research

Our toolbox is one of pragmatism and progress (that is, we not only believe in solutions, but hold them to be essential), typified by studies which are conducted *with* the community (collaboration and partnership), *for* the community (giving voice and ensuring everyone gains from insights and reaps the rewards) and *by* the community (citizen scientists and community member-led projects). In effect our work is interdisciplinary, multi-method and inherently critical: a diverse and flexible portfolio of what works, involving deductive and inductive techniques and data collection ranging from ethnography to statistics to content analysis to social network modeling. The overriding principle behind our research efforts is an ethical commitment to positive outcomes for the communities involved as well as individuals and our greater society as a whole. This drive rests on the assumption that the production of knowledge that happens in communities should help to drive the production of knowledge and systems of analysis or study present in universities. Further, most PAR adopters see research as subservient to community needs; if we walk away from a research project without significant or sufficient data but still leave the community better off than they were before, then we usually consider the effort to be a success. If you take this set of traits you find a tool set and perspective that's potentially independent of both information science and institutions. CI thus becomes more than just an emerging field, but a set of convictions, actions and ways of integrating ethics and agency into ones world view as an actor - be they a researcher, activist, policy-maker or in some other role.

Community Inquiry

Community inquiry presents an effective and appropriate informative model for CI. Professor Chip Bruce provides a compelling explanation:

"Community inquiry is inquiry conducted of, for, and by communities as living social organisms. Community emphasizes support for collaborative activity and for creating knowledge, which is connected to people's values, history, and lived experiences. Inquiry points to support for open-ended, democratic, participatory engagement. Community inquiry is thus a learning process that brings theory and action together in an experimental and critical manner." [4]

This definition features significant overlap with the PAR perspective presented above and draws upon John Dewey's rich conception of inquiry. It stresses addressing community-defined problems by building upon pre-existing local resources and knowledge and necessitates reflexivity - a questioning of community membership, values and goals - by representing the process as a cycle. This cycle is visually represented as a dynamic process of asking questions, performing investigations, creating understandings, and discussing and reflecting on them.

Cyberpower

As sites of public computing and potential places for community organization many libraries, civic centers and social service agencies have evolved in to what might be referred to as 'Community Technology Centers' or CTC's. Alkalimat and Williams [5] propose that CTC's may be a primary "organizational basis for democracy and social inclusion in the information society." [5]. Citing Tim Jordan [6], an STS researcher who was one of the first to critically pioneer the emerging landscape of culture, politics, power and inequality on the web and in information society at large, they explain that Cyberpower, "the effect of online activity on power" [5], is a potential measurable outcome from CTC's for individuals, groups and on an ideological basis. Cyberpower can be operationalized through a variety of metrics, such as valuable skills, experiences and accumulate social connections (in the case of Alkalimat and Williams, social capital), though always with a focus on an increase in an individual or group's ability to influence or address issues related to their needs. Ultimately Cyberpower suggests an emphasis on providing disempowered individuals more than just access to online activities and technology resources, but critical and creative perspectives that allow them to shape both the use of such tools and related behaviors and gain more control over their participation in our emerging information society [7]. This becomes a potential refutation of the critique that the employment of ICT's for development (ICT4D) is simply another project of digital capitalism and way to plug more poor people into consumerism and increase existing power disparities [8].

This trio of perspectives is a slice of the informative basis and interpretive framework behind methods present in community informatics. Our poster proceeds to present example CII projects related to each, which feature *integrative strategies* (storytelling with multimedia, relationship building, community memory, continuing education, and knowledge sharing), *future and current settings* (Sao Tome Africa, small town and rural CTC's, schools and libraries, and the local CU community), and *diverse audiences* (both children and adults, volunteers and CII staff, as well as community leaders).

Examples

We overview a set of Community Informatics projects, ranging from completed to in-planning stages, including:

Social and Environmental Justice On the Fifth and Hill Toxic Site

The problem is not new: a toxic site in the middle of a residential community, and an ongoing dispute between neighborhood activists and a large corporation about the health hazards it poses. The situation present in North Champaign is only a symptom of larger problems of social injustice related to race, health and corporate responsibility. It is no coincidence that the mostly African-Ameican neighborhoods are poor nor is it particularly surprising to find that the issue of environmental injustice sits alongside problems of poor relations with law enforcement, and lack of adequate support from local government. This project involved the use of mapping and new media communication technologies to present many of the environmental and health issues present in North Champaign.

Future Directions in Community Technology Center Research

Community technology centers, small libraries and non-profit organizations all struggle to manage their technology assets. Issues of digital literacy, external threats like computer viruses, rapidly changing hardware and burgeoning software options make today's IT environment difficult to navigate, even for experts. To meet this challenge, 21st century organizations require effective and robust management systems and education strategies that can deliver a variety of functions and positive outcomes. This semester long study focused on prototyping dynamic, web-based solutions for these challenges. They included:

- A dynamic inventorying system that tracks the 'health' of computers, and is capable of transmitting technical information to system administrators in the event of failure. IT administrators can access software and hardware information for each asset, quickly and easily.

- A geographic information system which helps visually organize the location of community technology centers in a city or region. GIS tools also enhance the overall situational awareness of organizations.
- A knowledge-sharing system to allow community help-desk organizations to disseminate critical information and improve training efforts. These type of systems help connect experts with beginners, and foster relationship building at all levels of the organization.
- Use of the Wordpress Content Management System (CMS) for truly community-driven web resources.
- Web-based technology training guides and tutorials designed to be modular, multimedia and most importantly, empowerment oriented. They not only teach essential digital literacy skills but also encourage relevant and critical use of technology through active and contextualized learning.
- A customized Linux install built specifically to support community technology education needs, which was combined with guides and documentation for ensured sustainability.

In addition we explored the possibilities of CTC-based education programs in a digital learning series for kids that included *Storytelling in [Stop] Motion* and *Comics and Community Stories*.

1. REFERENCES

[1] Gurstein, M. 2007. What is Community Informatics (and Why Does It Matter). *POLIMETRICA*.

[2] Williams, K., and Durrance, J. C. 2009. Community informatics. In *Encyclopedia of Library and Information Sciences*. Marcia J. Bates, Mary Niles Maack and Miriam Drake, editors. Taylor & Francis.

[3] http://www.cii.illinois.edu

[4] Bruce, B. 2008. What is Community Inquiry. *Chip's Journey: Thoughts About Community, Learning and Life*. Available online at http://chipbruce.wordpress.com/resources/community-inquiry-bibliography/what-is-community-inquiry/

[5] Alkalimat, A. 2004. Introduction to the Compilation and Social Capital and Cyberpower in the African American Community: A Case Study of a Community Technology Center in the Dual City. In *Cyberorganizing*. University of Toledo**,** available online at http://eblackstudies.org/grbk/

[6] Jordan, T. 1999. *Cyberpower: The Culture and Politics of Cyberspace and the Internet*. London: Routledge, Taylor & Francis Group.

[7] Banks, A. 2006. *Race, Rhetoric, and Technology*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

[8] Pieterse, J.N. 2005. Digital Capitalism and Development: the Unbearable Lightness of ICT4D in Geert Lovink and Soenke Zehle, Eds., *Incommunicado Reader*. Amsterdam: Institute of Network Cultures. Available online at http://networkcultures.org/wpmu/portal/publications/inc-readers/incommunicado/

Hogs and Harvesters in the Digital Age: The Farm, Field, and Fireside Collection at the UIUC Library

Harriet Elizabeth Green
University of Illinois at Urbana-Champaign
Graduate School of Library and Information Science
503 W Green St Apt 11
Urbana, IL 61801

The Farm, Field, and Fireside collection is a newly created and extensive digital archive of farm newspapers published in the Midwest from the mid---nineteenth century through the 1940s and 1950s. The wide array of publications in the collection------ranging from the hog farmer's Berkshire Stockman to the women's publication The Farmer's Wife------are an invaluable source of primary research materials and they offer a fascinating perspective on historical issues during a pivotal period of economic, social, and political growth in the United States.

The Farm, Field, and Fireside Collection is a stellar example of a digital library created to preserve a library's archived materials, and it employed digital tools for information organization, such as metadata to organize the collection and make it user accessible. The Farm, Field, and Fireside collection also is deeply immersed in the technologies of remembrance and forgetting, as it seeks to unearth and preserve a core part of Midwestern history.

This poster will discuss the creation and maintenance of the Farm, Field, and Fireside digital newspaper archive, and examine the goals behind its creation for the research needs of University of Illinois faculty, students, and staff, as well as the scholarly community at---large. The poster will explore: What research and information needs are addressed by the Farm, Field, and Fireside Collection? What was determined to be the most effective way of organizing the resources for students and researchers to use?

The poster will also review the topical web research guides that were created to support research conducted with the Farm, Field, and Fireside collection. This part of the poster will examine questions such as: What was the most effective and efficient format for delivering the research guides? How could we utilize the guides to connect the digital newspaper archive to the Library's large circulating and special collections? What type of background content was needed in the research guides in order to present the collection's items as viable sources for research? How often have the research guides been referred to by users while using the collection?

The poster session will explore how we addressed these questions by examining the maintenance and enhancement of the digital newspaper archive with Olive Active Paper Librarian software, and explaining the process of creating and maintaining web research guides that promote the Collection's interdisciplinary connections to the Library's broad print and digital collections. This poster will offer an intriguing examination of a new and wholly unique digital archive that is a valuable resource for historians, digital humanists, students, information professionals, and all interested public citizens.

The need for qualitative methods in online user research in a digital library environment

Elke Greifeneder
Berlin School of Library and Information Science, Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin, Germany
+49-30-2093-4494
greifeneder@ibi.hu-berlin.de

ABSTRACT
Online users of digital libraries are multi-local, multi-lingual and live in multiple time-zones. Getting "purposeful data" in online user research requires that the research be done online because the users are there. This content analysis looks at a broad sample of international publications to address the following two research questions: 1) what methods do we use for online user research and 2) what are the purposes behind the research questions? The poster suggests that we currently use methods that match poorly to the purpose of the study and that there is a real need to use qualitative methods to study online users to be able to produce purposeful data.

Categories and Subject Descriptors
H.3.7 [Digital Libraries] : User Issues

General Terms
Measurement, Human Factors

Keywords
Digital Library, Qualitative Methods, User Research, Research Design

1. INTRODUCTION
Online users of digital libraries are multi-local, multi-lingual and live in multiple-time-zones. If we build up a digital library based on results from a focus group with local users, the results may not represent the real future users. In studying online usage, we can no longer rely only on local users. Getting "purposeful data" (in Troll Covey's sense of the word) in online user research requires that the research be done online because the users are there. Not all methods are currently used online. Focus groups can be difficult to do online, as are interviews and ethnographic observations. Structured surveys and log file analyses are, on the other hand, widely used for online studies. The proposed poster presents a content analysis of user research drawn from published articles and dissertations revealing that log file analyses and surveys are the primary methods used for online research.

2. RELATED WORK
Most of the research on users and digital libraries analyzes one digital library and its users. The number of publications on specific methods is limited. Edgar [1], Homewood [2], Xia [3], and Nicholson [4] discuss general research design issues and do not map methods to purposes. Studies like "The virtual scholar: the hard and evidential truth" [5] try to draw a general picture of online users. The poster's content analysis grows out of Troll Covey's study [6] that addresses the relation of purposes and methods in online user research. She undertook interviews with participants from the Digital Library Federation (DLF) about their use of and experience with methods in user research and concluded that "Libraries are struggling to find the right measures on which to base their decisions. DLF respondents expressed concern that data are being gathered for historical reasons or because they are easy to gather, rather than because they serve useful, articulated purposes."

3. RESEARCH DESIGN
Purposeful data results from an expressed purpose in combination with an adequate method. Data gathering is an essential part of online user studies, and every method has its areas of application and its limitations: quantitative surveys are limited in their ability to detect causal relations; with qualitative interviews broad generalizations are risky. This poster looks at a broad sample of international publications to address the following two research questions: 1) what methods do we use for online user research and 2) what are the purposes behind the research question? The content analysis follows the thematic coding in Hopf [7]. This study extends Troll Covey's work by correlating the findings with new variables such as whether the research has taken place offline or online and whether the result fits the purpose. The 70 publications that have been taken into consideration contain applied user research in a digital library environment and examine only online services. The databases DABI, E-LIS, DissOnline, ProQuest and LISA served as the sources.

4. RESULT ANALYSIS

Because of space considerations, only one result from the analysis will be presented in the poster. Only in 21% of the cases did researchers use a qualitative method (focus groups, interviews or observations) – more than double that number used surveys. Methods like interviews or focus groups currently take place nearly entirely offline. Only one of the studies used an online interview. Qualitative methods traditionally require human-to-human communication – for example the interviewer and the interviewee – to be able to reformulate a question or to respond to a specific answer in order to get deeper insights into behavior. An example of qualitative research in usability engineering is the construction of personas and scenarios for a digital library – for both, deep insight into sample users is needed, not the whole population. Quantitative methods may be used afterwards to check if the personas or the scenarios match the population. Despite artificial intelligence experiments, machines currently cannot be programmed to conduct unstructured interviews on their own – for example a chat system always needs a human behind the machine. The problem is that quantitative research designs require knowledge about the user's context to be able to ask the right questions and to interpret the data in the right way. Do closed answer-sets offer the options that users would provide or only the questioner's perspective? Can log files be analyzed without knowing the full social context of the users' actions? If most people select new offerings, does this mean that they want that particular information or are merely browsing? The poster's analysis shows that surveys are used for many more purposes than all the other methods and that they are used as an all purpose research tool for need assessments, user typologies, perception studies, satisfaction testing – even testing usability.

5. CONCLUSION

Although researchers may use quantitative methods, they tend to articulate purposes like user typologies or need assessments that implicitly demand qualitative methods with an interactive human presence. If the purpose is to know users and the context in which they use a digital library, human-mediated inquiries need to substitute surveys and log file analysis. As Notess says: "Part of the problem is that the log files do not tell us anything about user motivation or rationale. For instance, we noted that only 11% of user sessions used bookmarking. But we do not know why the other 89% did not make use of this feature."[8] The poster suggests that current methods tend not to match the intended purpose of user studies study and that there is a real need for qualitative data in online user research.

6. REFERENCES

[1] Edgar, B. 2006. Questioning LibQUAL+: Critiquing its Assessment of Academic Library Effectiveness. In Proceedings of the ASIS&T Annual Meeting. DOI= http://www.asis.org/Conferences/AM06/papers/112.html

[2] Homewood, J., Huntington, P.and Nicholas, D. 2003. Assessing used content across five digital health information services using transaction log files. Journal of Information Science 29 (6 2003)

[3] Xia, W. 2003. Digital library services: perceptions and expectations of user communities and librarians in a New Zealand academic library. Australian Academic and Research Libraries 34 (1 2003), 56-70. DOI= http://www.alia.org.au/publishing/aarl/34.1/full.text/xia.html

[4] Nicholson, S. 2005. A framework for Internet archeology: Discovering use patterns in digital library and Web–based information resources. First Monday 10 (2 2005). DOI= http://firstmonday.org

[5] Nicholas, D. 2010. The virtual scholar: the hard and evidential truth. In Digital Library Futures. IFLA Publication Series. K.G. Saur Verlag, Munich, in print

[6] Troll Covey, D. 2002. Usage and Usability Assessment: Library Practices and Concerns. Council on Library and Information Resources, Washington, DC. DOI= http://www.diglib.org/pubs/dlf096/dlf096.html

[7] Hopf, C. and Weingarten, E. 1984. Qualitative Sozialforschung. Stuttgart: Klett-Cotta

[8] Notess, M. 2006. Three looks at users: a comparison of methods for studying digital library use. Information Research 9 (3 2004)

Achieving the Intercalation of the Social and the Technical in Computing: The SREC (Socially Robust and Enduring Computing) Program

Vincenzo D'Andrea
Faculty of Sociology and Department of Computer Science,
University of Treno
+39 0461 88 2084
vincenzo.dandrea@unitn.it

David Hakken
School of Informatics
Indiiana University
Bloomington, Indiana, USA
+1 812 856 1869
dhakken@indiana.edu

Maurizio Teli
Trento Museum of Natural Science
Via Calapina
14Trento, Italy
+39 0461 270332
maurizio.teli@maurizioteli.eu

ABSTRACT
This Poster addresses the core issue of iSchool research: How to achieve adequate accounts of what happens when people and organizations use automated (that is, computer-based) information and communication technologies. It outlines the authors' vision of how persuasive, balanced accounts that properly integrate social and technical perspectives on computing might be achieved, and a research program intended to make this vision real.

Categories and Subject Descriptors
K. Computing Milieux. K.0 Computing Milieux, General K.7.0 [**The Computing Professions, General**]

General Terms
Management, Performance, Design, Economics, Reliability, Human Factors, Standardization, Theory, Legal Aspects.

Keywords
Combining Technical and Social Aspects of Computing, Professional Institutionalization of Computing, the Future of Computing Keywords are your own designated keywords.

1. INTRODUCTION
Computer-based information systems (CISs) appear in an ever increasing array of human activities, and these systems generally continue to grow in complexity. The humans who try to use the information CISs produce continue to have problems with them.

While doubtless some of these problems encountered are new and unique to computerizing new domains, others are familiar, similar to problems encountered before—e.g., a new document-handling system not well articulated with legacy systems, so things one could do with the old system, like scanning, are no longer possible. This poster describes several elements of a program to eliminate at least some of these continuing problems and thus aims to chart a way forward for computing.

2. EMPIRICAL BASE
Our project is based first on a reading of the substantial body of empirical research on and professional practice in computing, much of it fostered in iSchool environments. Our reading of this literature stresses how many of computing's problems derive from failing to make CISs socially robust enough. That is, they don't endure because of inadequate or improper attention to the social in one or more elements of the melange of design, implementation, and/or maintenance. This research suggests that greatly increasing their social robustness could mean significant improvement in how they work.

3. THE ENSUING ANALYSIS
In our view, the failure to make computer-based information systems more socially robust and therefore more enduring is not a matter of oversight. Rather, it is connected to important aspects of how computing, in both disciplinary and "in practice" aspects, has come to be socially constructed. An almost exclusive focus on technical virtuosity, for example, has obscured the need for social virtuosity.

Given this professional history, computer-based information systems (CISs) will not become socially robust merely by recognizing that they are not and wishing them to become so; additional steps are necessary. In the poster, we illustrate these points through a brief characterization of earlier as well as more recent approaches to software development and via discussion of earlier attempts within computing to take account of the social. This is followed by discussion of some recent forms of computing, notably Free/Libre and/or Open Source Software development projects, "Web 2.0" practices, and Participatory Design. We content that it is not enough, as systems in these

forms do, to alternate some social with the predominantly technical moments. Even these quasi-socially robust forms of computing, while suggestive, remain insufficiently robust socially for reasons ultimately very similar to those hampering other CISs. In sum, analytically, our program identifies several barriers internal to the current social construction of computing disciplines and professions which interfere with attaining more properly balanced CISs and therefore with making them more enduring.

4. THE PROGRAMATIC IMPLICATION: INTERCALATION OF THE SOCIAL AND THE TECHNICAL AS THE GOAL

Rather, basic approaches to CISs need to be *re*constructed if they are to become socially robust enough. To illustrate how this might be done, we present a conception of computing in which the social and the technical are effectively intercalated. This conception provides a vision powerful enough to base sufficiently socially robust computing. We draw on the Science, Technology, and Society literature to conceptualize a possible future relationship between social and the technical moments in these disciplines, that of "intercalation" [1]. In a state of intercalation, the various moments that make up a profession operate in mutual respect but often at some distance from each other. We argue that, as it would create enough conceptual space for more enduring information systems, intercalation of the social and the technical is the most plausible and yet still worthwhile condition for the computing disciplines to aim at.

5. THE RESEARCH PROJECTS

Finally, the poster outlines key elements of our research projects on Socially Robust and Enduring Computing. SREC involves both field and action research on the theoretical and practical problems of integrating social and technical perspectives in existing projects. It aims to illuminate the prospects for intercalation of the social and technical by testing whether such a relationship indeed increases the ease of implementation, the utility, and the longevity of computer-based information systems. (The SREC program and its research projects are an international collaboration, being carried out by faculty from the Faculty of Sociology and the Department of Computer Science and Engineering at the University of Trento in Italy, faculty from the Social Informatics group in the School of Informatics at Indiana University/Bloomington, and other associated scholars.)

6. CONCLUSION: THE SREC PROGRAM AND THE INTERCALATION OF THE SOCIAL AND THE TECHNICAL IN COMPUTING

In the absence of shared understandings of how to intercalate the technical and the social, we expect humans to continue to have difficultly making CISs work well. Achieving such understandings depends upon not only acknowledging computing's sociality but also on embodying them into the development process.

The SREC program is intended to be an effective test of whether our envisioned reconstructive intervention is likely to work. Its foundational goal is to develop more inclusive intellectual principles on which to base computing practices. This goal will be achieved, we believe, by organizing software development to be properly symmetrical regarding its social and technical aspects. As it is being carried out collaboratively by scholars from both the computing and the social sciences, SREC combines methodological approaches used in computer science (such as creation and testing through use of development tools) with those in social science, including ethnographic studies of "quasi-socially robust" computing. Further, it is an important goal of SREC to prefigure and indeed to illustrate what computing practices that properly integrate the social and the technical would look like, what this would mean in practice. For example, SREC's applied research program embodies these understandings in design specifications that are then built into new design tools, as well as working with local firms and organizations to develop their own SREC capabilities.

7. ACKNOWLEDGEMENTS

Our thanks to our colleagues and interlocutors in Trento and Bloomington, including those who have been supportive of the developing relationship between our universities and our research programs. An initial element of the SREC program has been approved for funding by the Province of Trento, Italy; actual release of funds is still pending.

8. REFERENCE

[1] Galison, P. 1997 Image and Logic. University of Chicago Press.

Fighting Diabetes with Information: Where Social Informatics Meets Health Informatics

Barbara Hayes
Indiana University School of Informatics at Indianapolis
535 W. Michigan Suite 475B
1-317-278-7672
bmhayes@iupui.edu

William Aspray
School of Information
University of Texas at Austin
D8600, Austin TX 78701
1-512-471-3877
bill@ischool.utexas.edu

ABSTRACT
This abstract sets out a research agenda for information scientists and technologists interested in the interrelationships among patients, health care providers, and information technology. Using the complex and costly diagnosis of diabetes as a vehicle for exploration, this work suggests addressing a set of problems that will improve the lives of patients, their families, and friends, as well as making the provision of diabetes care more effective and cost efficient. Information technology tools and methods are used, but with sensitivity to the social and organization complexities of health care. I-Schools graduates, with their interdisciplinary mindset, social science methodologies, and familiarity with IT and its applications can increase the success rate of IT interventions in health care. Topics include public health and community informatics, knowledge dissemination, information alerts, decision support, clinical guidelines, health literacy, patient, pharmacy, and laboratory feedback systems, interface design, reminder systems, consumer informatics, and privacy and security issues.

Categories and Subject Descriptors
J.3 [Life and Medical Science]: Health, Medical Information Systems

General Terms
Management, Documentation, Design, Economics, Reliability, Security, Human Factors, Legal Aspects

Keywords
Information research, diabetes, informatics research, public health and community informatics, knowledge dissemination, management alerts, information alerts, decision support, health literacy, technological literacy, patient feedback systems, pharmacy feedback systems, laboratory feedback systems.

1. INTRODUCTION
Careful study of the complex interrelationships among people, information, and technology holds great promise for improving American health care. Health care professionals have attempted to use various forms of information technology (IT) to improve the fragmented health delivery system for decades, but financial resources, technical expertise, and interdisciplinary research by information and health researchers have been scarce. The Obama administration has decided to infuse billions into health information technology (HIT). Here, finally, is a chance to "jumpstart" the process of improvement. But do information researchers really know how to spend this money? Will they deploy it effectively? Are technologists and health care professionals actually ready to join forces to create effective new solutions for patients?

This research examines several information challenges associated with Type 2 diabetes to allow information researchers unfamiliar with healthcare to observe the social and organizational factors in the ebb and flow of information around complex diagnoses. Diabetes is a widespread, debilitating and expensive disease. The Centers for Disease Control and Prevention estimated the prevalence of diagnosed and undiagnosed diabetes in the United States in 2007 at all ages was 23.6 million people or 7.8% of the population. The Center further estimates that at least 57 million American adults had prediabetes in 2007 (Centers for Disease Control and Prevention 2007). Prediabetes is a condition in which individuals have blood glucose levels that are higher than normal but not yet high enough to be diagnosed as diabetes (American Diabetes Association 2007).

The American Diabetes Association reports that $174 billion was spent on diabetes in 2007, which is $42 billion more than was spent in 2002 (Berger 2007; American Diabetes Association 2008). The indirect costs of diabetes due to reduced performance, lost productivity, early mortality, and disability are estimated at close to $58 billion (American Diabetes Association, 2008a; Berger, 2007). Diabetes also takes a high toll on individuals and families. Medical expenses of patients diagnosed with diabetes are 2.3% higher than those who do not have the disease (American Diabetes Association, 2008). In 2007, the per capita annual health care costs for people with diabetes were estimated at $11,744 per year (American Diabetes Association, 2008).

The management of diabetes requires a partnership between health care providers and patients that is largely driven by information needs. Health care providers, patients, and families must all be armed with adequate information to attenuate risk factors, make the almost daily essential adjustments in diet, exercise, medications, and manage complications that are required to return people with diabetes to good health. While major advances are being made in metabolic and pharmacologic science,

there has been no systematic assessment of patients' information and technology needs and how to go about addressing them. This abstract focuses on some of the initial efforts that information scientists and technologists have made to help patients with diabetes, as well as a resulting research agenda that will soon be published in collected form. An examination of diabetes will allow information scientists to explore the use of information technology as a tool to manage any kind of chronic illness.

Methods

Utilizing literature review, the authors identified five domains of the disease that lend themselves to the identification and provision of important information. These areas of inquiry are: (1) risk assessment and mitigation (2) pre-diabetes (3) receiving and assimilating a new diagnosis of diabetes (4) understanding, preventing, and managing complications of diabetes and (5) end-stage disease

Medical, nursing, information science and technology researchers working in the area of diabetes were then identified and asked to discuss the complex interrelationships among genetic and environmental factors; patient characteristics; health care provider knowledge and behaviors; and the information and technology needed to manage these complexities to lead to better health outcomes for persons with diabetes. These discussions provide a robust framework for an information research and teaching agenda to address these critical information needs.

Results

The research and development agenda emanating from this work is organized into six sections of an edited book currently in press (Hayes and Aspray 2010). The work begins with an overview of improving all aspects of diabetes care through information technology. Ensuing chapters examine early efforts to create and adapt technologies to improve diabetes care. They are organized into sections on ubiquitous computing; using educational gaming to educate and treat diabetes; other technological explorations of diabetes care; using technology to improve patient access to information; and methodological and theoretical considerations.

The last chapter of the book examines these kinds of collaborations, their challenges, and ways to foster them. The disciplines of health care and technology differ greatly in their approaches to problems, training, and publishing. Health care professionals are used to "running the show." They have extensive content knowledge. Information technologists can gauge how difficult it may be to build and deploy useful tools and they have more knowledge of usability as it applies to digital tools. Health care providers, technologists, and patients who have firsthand knowledge of the disease must all come together at the drawing board in an interdisciplinary, collaborative way.

Using the complications of diabetes for illustration, it becomes apparent that there are many ways in which technology may be applied to help patients cope with disease burden. The following table provides examples:

Information Need	Technology Intervention
Information exchange/sharing among multiple specialists	"Hand-off" systems incorporated into digital electronic records
Education for patients dealing with ophthalmologic, cardiovascular, neurologic complications of diabetes	Assistive information technologies built to accommodate poor eyesight, fatigue, and decreased feeling in the fingertips
Barriers to visiting providers because of increasing infirmity or episodes of illness	Secure communication portals and mobile devices that can upload and assess data on patient status
Coordinating care of patients taking multiple medications and visiting multiple healthcare providers	Decision support systems that can give visual "snapshots" of patients' recent care and advise generalist physicians on specialty topics
Ways to monitor frail patients in their homes	Sensors/ubiquitous computing to provide in-home monitoring

Significance

Information technology has had an impact on health care, but only at the most superficial level. IT is used primarily to document episodes of care and secure payment for care. It has often fallen short as a tool to support medical decision making and improve the lives of patients and their families (Committee on Engaging the Computer Science Research Community in Health Care Informatics 2009). Opportunities abound to build transformative, intelligent information systems that meet the needs of sick people. At the same time, we risk squandering huge sums of money on poorly designed systems that fail to meet those needs because they underestimate the complex social and organizational health care milieu. This book is a beginning effort to inform those efforts.

References

American Diabetes Association (2007). "How to Tell If You Have Pre-Diabetes." from http://www.diabetes.org/pre-diabetes/pre-diabetes-symptoms.jsp.

American Diabetes Association (2008). "Economic cost of diabetes in the U.S. in 2007." Diabetes Care 31(3): 596-615.

Berger, J. (2007). "Economic and clinical impact of innovative pharmacy benefit designs in the management of diabetes pharmacotherapy." American Journal of Managed Care 13 (Suppl2),: S55-S58.

Centers for Disease Control and Prevention (2007) National diabetes fact sheet: general information and national estimates on diabetes in the United States.

Committee on Engaging the Computer Science Research Community in Health Care Informatics (2009). Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. Washington, D.C., The National Academies Press.

Hayes, B. and W. Aspray, Eds. (2010). Health Informatics: A Patient-Centered Approach to Diabetes. Cambridge, Massachusetts, MIT Press.

Automated Keyword Extraction of Learning Materials Using Semantic Relations

1. INTRODUCTION

The poster will present our on-going research, which will develop new algorithms to automatically generate keywords from online documents that describes lesson plans in mathematics and science. The motivations for improving the current keyword extraction mechanism are twofold:

- Feedback from our previous study (described below) showed that the keyword extraction was the least satisfying component of our automatic metadata extraction mechanisms to the users.
- Our data indicated that human annotators often assigned keywords to a document that do not appear in the document, which were impossible for the current keyword extraction mechanism to generate.

Building upon TextRank by Mihalcea and Tarau [4], our approach is to use a graph-based algorithm to rank keywords, based on semantic relationships.

2. BACKGROUND

A team comprised of the Center for Natural Language Processing (CNLP) at Syracuse University and the Digital Learning Sciences (DLS) at the University Corporation for Atmospheric Research recently completed a project that integrated many digital library tools into one, which is called Metadata Assignment and Search Tool (MAST). [1] This tool enables libraries and museums to efficiently describe and disseminate their digital materials by 1) automatically generating metadata to assist the cataloger; 2) assisting in assigning educational standards to learning materials; and 3) customizing their workflows and collection management. Previous versions of these tools are deployed in the National Science Digital Library (NSDL) project to assist catalogers in adding materials to the online digital collection.

[1]The project is funded by the Institute of Museum and Library Services (IMLS).

The automatic metadata assignment uses Natural Language Processing technologies to process the text of the online documents, in html or pdf formats, and produces metadata elements for the Dublin Core + GEM fields. These fields include general elements such as title, description, subject fields and contributors, as well as educational fields such as audience, instructional methods and grade level. The automatically generated fields are presented to the collection cataloger, who may correct or add to them. In user study tests with a group of managers, curators and directors representing both museums and libraries, their reviews of the process of cataloging with automatic metadata suggestion and managed workflows were generally favorable and enthusiastic. However, part of the feedback received from this group was that while almost all of the automatic metadata was helpful in cataloging, subject terms were not. Thus our current research lies in improving the automatic generation of these subject terms, which we will call keyword extraction, in keeping with the terminology used by other researchers in this area.

3. RELATED WORK

The idea of using the graph-based approach for information retrieval systems appeared in the early days of the research, such as THOMAS, a human-machine dialogue system, by Oddy [5]. An explicit use of the approach, however, was first introduced by Preece [6], where a mechanism called *spreading activation* was employed. Spreading activation is an algorithm for searching networks, which starts from a single node and spreads out to other nodes through edges while assigning a weight (or "activation") to each node. Although various work has been done using the spreading activation algorithm since then, it was the success of PageRank [1], which demonstrated the usefulness of the approach to the research community as well to the general public. The idea of PageRank is to utilize the hyperlink structure of Web documents, in addition to their contents, to rank retrieved documents. PageRank constructs a graph structure, where hyperlinks are represented as edges, and web documents are represented as nodes. It then assigns higher weights to nodes with more edges coming from other nodes. Thus, it effectively collects "votes" from Web pages to rank the pages.

The strength of the graph-based approach is the generalizability: the approach may be applied to any structural relations, not just the hyperlinks. Mihalcea and Tarau proposed TextRank [4], a graph-based ranking algorithm similar to PageRank, for extracting keywords from documents.

With TextRank, nodes in the graph represent words instead of Web documents, and edges represent word co-occurrences instead of hyperlinks. Coursey et al. applied the TextRank algorithm to extract keywords from learning materials of history, combined with another keyword extraction method called Wikifier [2].

4. CURRENT STUDY

4.1 Approach

Our approach is to enhance TextRank by using semantic relations, instead of term co-occurrences, as edges of the graph. Specifically, we plan to apply the definition of "semantic relatedness" using Wikipedia by Strube and Ponzetto [7]. While previous definitions of semantic relatedness have been based on word relations derived from sources such as WordNet [3], more recently researchers have used the resources available through Wikipedia, as in [7]. As a collaboratively generated corpus, Wikipedia provides a breadth and depth of topics not easily achieved otherwise. In order to define semantic relatedness, Wikipedia pages can be viewed as a collection of categorized concepts, forming a semantic network. Relations between concepts are given by a hyperlink structure between articles, forming a wide variety of relations, not just "is-a" or "part-of" relationships.

4.2 Evaluation Environment

A platform for evaluating and developing keyword extraction mechanisms has been developed using Java and CNLP's libraries that utilize the text processing engine, TextTagger, and other utilities such as html or pdf document preprocessors. The evaluation environment is depicted in Figure 1. For each learning material, the text is processed with TextTagger and the accompanying preprocessing to obtain MAST metadata. The different keyword extraction algorithms, shown as 1 through N, may use the the extracted metadata, the text directly from the learning materials and statistics from a background corpus in their definitions. These algorithms will include a baseline standard keyword algorithm, Mihalcea's TextRank algorithm, and our new algorithm based on TextRank with semantic relations. The evaluation module will compare the different keyword extraction algorithms by calculateing standard measures (precision, recall, and f-measures) based on the gold standards, which are extracted from manually created MAST metadata.

4.3 Data

The study utilizes metadata that have been created in the MAST project (described in the Background section), as the gold standard. Human annotators (information professionals) created a list of keywords that describe the document as part of the metadata. So far metadata were created for 50 online documents, which described lesson plans in mathematics and science.

5. FUTURE WORK

At the time of writing, we are implementing two keyword extraction algorithms: TextRank with word co-occurrences and TextRank with semantic relatedness. By the time of the conference, we aim to test the two algorithms against the data and present the results in the poster.

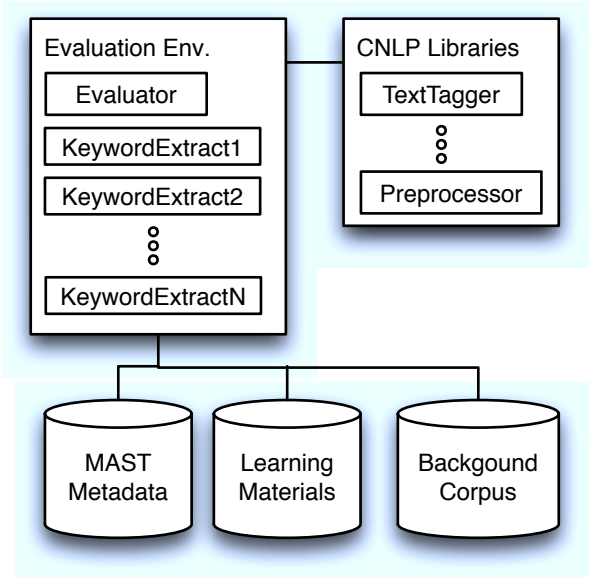


Figure 1: Evaluation Environment

6. REFERENCES

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 1998.

[2] K. Coursey, R. Mihalcea, and W. Moen. Automatic keyword extraction for learning object repositories. In *Proceedings of the Conference of the American Society for Information Science and Technology*, Columbus, Ohio, October 2008.

[3] C. Fellbaum, editor. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA, 1998.

[4] R. Mihalcea and P. Tarau. Textrank: Bringing order into texts. In *in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, Barcelona, Spain, July 2004.

[5] R. N. Oddy. Information retrieval through man-machine dialogue. *Journal of Documentation*, 33(1):1–14, 1977.

[6] S. E. Preece. *A spreading activation network model for information retrieval*. PhD thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 1981.

[7] M. Strube and S. P. Ponzetto. Wikirelate! computing semantic relatedness using wikipedia. In *The Proceedings of the Twenty-First National Conference on Artificial Intelligence*. American Association for Artificial Intelligence, July 2006.

A Socio-Technical Analysis of the Interplay between Inter-Organizational Information Technologies and the Network Forms of Inter-Organizational Governance

Mohammad Hossein Jarrahi
The School of Information Studies
Syracuse University
mhjarrah@syr.edu

ABSTRACT

Due to rising collaborations among organizations and permeation of various ICT, Inter-organizational information systems (IOS) continue to attract scholarly attentions. The use of IOS technologies enables organizations to engage with other organizations in ways that previously seemed impossible and to present unprecedented inter-organizational arrangements. Despite their importance, little research has been conducted to conceptualize the mutual relationship between IOS technologies and new inter-organizational arrangements that are reflected in inter-organizational governance structures. In this poster, I will identify an emerging gap within the relevant literature. Then, a theoretical foundation will be presented by way of this research problem. Finally, I will discuss possible strategies to empirically investigate the research problem.

1. MOTIVATION

The extensive and dynamic business operations conducted by many organizations require an ability to extend external ties and undertake inter-organizational collaborations. At their most basic level, inter-organizational information systems are shared by two or more organizations, and are designed to link business processes [2, 3]. More generally inter-organizational information systems denote the uses of ICT that transcend formal organizational boundaries to provide both structures and flow of information and knowledge among players [4]. IOS technological infrastructures include, but are not limited to Electronic Data Interchange systems (EDI), the Extranets, and the Internet.

A number of studies have investigated the influences of IOS technologies on the inter-organizational structures, and the way organizations conduct inter-organizational transactions. However, Most of these studies adopted a technologically deterministic approach which perceives the IOS technology independent of its organizational and social contexts [5]. This view tends to mask the mutual interdependency of organizations and technologies. [6].

In addition, the limited amount of research which has investigated the broader implications of IOS technological infrastructure has mostly adopted economic or strategic lenses [7]. The dearth of research that leverages other theoretical perspectives, particularly context-aware theories, can obscure our understanding of the relationship between the adoption of IOS technologies and inter-organizational governance structures.

To address the IOS governance, it stands to reason to draw on the network perspective which recognizes the embedded nature of social relations [8]. Certainly, more research is needed to study the mutual constitution of network forms of organization (as an alternative to stylized hierarchy and market) and new IOS technologies that indicates different organizational and social consequences than those of hierarchy-based EDI systems. The contextual insights afforded by network perspective can complement economic perspective and explain how and why technologies are used in practice.

Therefore, the main aim is to conduct an empirical research which seeks to conceptualize the interplay between IOS technological infrastructure and the network forms of inter-organizational governance. To do so, I would draw on socio-technical theories which acknowledge the situated entanglement between technology and social orders, and provide plausible means to account for governance structures.

2. THEORETICAL FOUNDATIONS

Since the subjects of my study—IOS technologies and their effects on IOS governance —are a multi-faceted phenomenon and spans multiple levels of analysis, it will be necessary to adopt multiple research perspectives and theories when constructing a conceptual framework. In this section, I will introduce two sets of theories and illustrate how they can contribute to the understanding of the critical aspects of my research problems.

2.1 Actor Network Theory

Conceptual insights from Actor Network Theory(ANT), can address the first part of the problem, and provide

balanced insights into how IOS technologies and human actors are mutually and emergently productive of one another [6]. ANT theorizes technological networks as embracing not only the human actors but also the physical artifacts and the concepts to which those actors relate [e.g., 9, 10]. At the heart of ANT lies the concept of generalized symmetry which implies that all the heterogeneous elements of a network, both human and non-human, can be explained in the same terms.

Breaking away from social and technological deterministic views, ANT provides precedents for understanding the contribution of both humans and artifacts to the innovation processes. It explains how certain technology constructs the identity of the other actors by making the latter act in accordance with its wishes. Therefore, ANT does not assume an *a priori* relationship between the social and the technical. Latour [10] notes that they can only be understood as inseparable and situated relationships between various human and non-human actors. Therefore, based on the ANT conception, IOS technology should not be considered a set of tools to be used to accomplish some tasks, but are constitutive of both practices and identities.

2.2 Institutional Theories

The Scott's [11] layered model of institutions and their environments can account for the second pillar of the research problem which is concerned with the inter-organizational governance structure. This model has a strong resonance with the network perspective.

Scott and Davis [11] - inspired by open-system theory- propose a multi-layer model which directs attentions to events and processes external to organizations [See Figure1]. Among proposed levels, organizational populations, and organizational field are deemed relevant to my study. Organization populations are defined as aggregates of organizations that are similar in some respect. They are clusters of organizations that produce similar products and survive, operate in similar institutional environments and shared the same normative, cognitive and regulatory structures [12]. Organization fields are collections of diverse types of organizations engage in competitive and cooperative relations. The notion of organizations fields essentially exhibits a higher level of environments than organization populations, and denotes structures that are collectively beneficial, improving adoption to the environment for all its members.

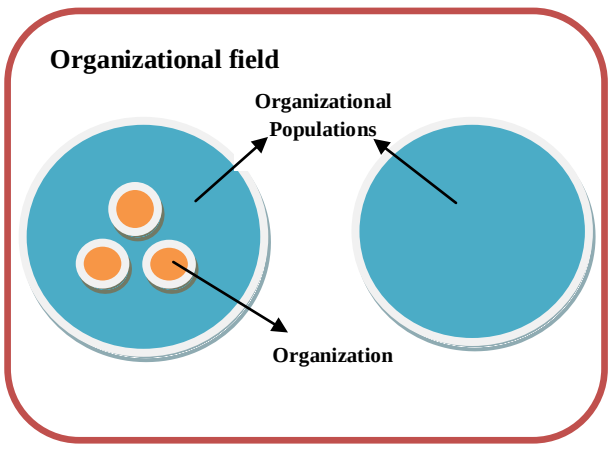


Figure 1: organizational environment (Adapted from [1])

ANT is argued to be sensitive to more micro interactions [13]. As such, the handicap of ANT (regarding broader consequences) can be offset by more macro analyses afforded by Scott's model, which can incorporate IOS effects at inter-organizational level. Finally, institutional view on environmental forces can explain how organizational and inter-organizational contexts shape the uses of IOS artifacts in lower levels. To this end, it relates the organizations or industry structures to ongoing actions of social actors [14].

3. SITES SELECTION

There are different plausible ways to select the cases for my proposed study. The following table showcases the disparate combinations based on the two major elements of the research question.

Table 1. Different site selection strategies

		Governance structure	
		Same	**Different**
IOS Technological infrastructure	**Same**	**1**	**2**
	Different	**3**	**4**

Both option 2 and option 3 look viable for addressing the research questions. Barley [15] has used strategy 2, and studied how an identical CT scanner technology has produced structuring processes and differing forms of organizations in two radiology departments. However, I argue that option 3 could serve my research question better. Leonardi and Barley [16] contend that since the mid-1980s researchers have adopted research designs that compare the use of identical or similar technologies in different context to highlight the role that social context play in shaping the technological consequences. They argue that the agenda of socio-technical researchers should instead turn into the opposite approach: comparing radically different technologies in the same or similar context.

To this end, I would like to study different IOS technologies in *similar* contexts of inter-organizational arrangement that are compatible with our definition of the network form of governance. Within such settings, long-term relationships are sustained while no central point of authority governs them.

This Multiple-case design should include multiple network-oriented inter-organizational settings that share a great deal of commonalties. The cases should also include a

constellation of organizations as IOS governance typically includes more than one organization. As Barley asserts: "Not only are organizations suspended in multiple, complex, and overlapping webs of relations, but the webs are likely to exhibit structural patterns that are invisible from the perspective of a single organizations caught in the tangle. To detect overarching structure, one has to rise above the individual firm and analyze the system as a whole." [17 , p. 321]

4. CONCLUSION

I argue that little is known about the nature and inherent influences of IOS technologies on the inter-organizational governance structure. Hence, explicit considerations of inter-organizational structures and their possible interactions with the IOS artifacts could heighten our understanding about the social and the technical aspects of inter-organizational information systems.

The theoretical frameworks for studying the research problem seem insightful enough to capture the relationships between macro structures and micro dynamics. Through the concept of symmetry, ANT recognizes the fact that IOS technologies are both the products and the shapers of human actions. In this light, our main argument is that viewing technology as fixed artifacts that are to be distributed throughout society impedes understanding, and therefore managing, the often necessary process of mutual adaptation of the social and the technical. In addition, the institutional view directs attentions to macro structures and processes, and hence complements ANT by integrating broader levels of analysis.

5. REFERENCES

[1] M. W. Chiasson, and E. Davidson, "Taking industry seriously in information systems research," *Mis Quarterly*, vol. 29, no. 4, pp. 591-605, 2005.

[2] J. Y. Bakos, "A strategic analysis of electronic marketplaces," *Mis Quarterly*, pp. 295-310, 1991.

[3] J. I. Cash, and B. R. Konsynski, "IS redraws competitive boundaries," *Harvard Business Review*, vol. 63, no. 2, pp. 134-142, 1985.

[4] M. L. Markus, C. W. Steinfield, R. T. Wigand *et al.*, "Industry-wide IS Standardization as Collective Action: The Case of the US Residential Mortgage Industry," *Mis Quarterly*, vol. 30, no. 2, pp. 439–465, 2006.

[5] D. Robey, G. Im, and J. D. Wareham, "Theoretical Foundations of Empirical Research on Interorganizational Systems: Assessing Past Contributions and Guiding Future Directions," *Journal of the Association for Information Systems*, vol. 9, pp. 497-518, 2008.

[6] W. J. Orlikowski, and S. Scott, " Sociomateriality: Challenging the Separation of Technology, Work and Organization " *The Academy of Management Annals*, vol. 2, pp. 433-474, 2008.

[7] K. Crowston, and M. D. Myers, "Information technology and the transformation of industries: three research perspectives," *Journal of strategic information systems*, vol. 13, no. 1, pp. 5-28, 2004.

[8] J. M. Podolny, and K. L. Page, "Network forms of organization," *Annual review of sociology*, vol. 24, no. 1, pp. 57-76, 1998.

[9] M. Callon, "Actor-network theory: the market test," *Actor network theory and after*, pp. 181-195, 1999.

[10] B. Latour, *Reassembling the social: an introduction to actor-network-theory*: Oxford University Press, USA, 2005.

[11] W. R. Scott, and G. F. Davis, *Organizations and organizing: Rational, natural, and open system perspectives*: Pearson Prentice Hall, 2007.

[12] W. R. Scott, *Institutions and organizations*: Sage, 2001.

[13] D. A. Howcroft, and N. N. Mitev, "M. Wilson (2004),'What we May Learn from the Social Shaping of Technology Approach,'" *Social Theory and Philosophy for Information Systems*, pp. 329-371, 2004.

[14] A. Giddens, *The constitution of society*: Polity Press Cambridge [Eng, 1984].

[15] S. R. Barley, "Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments," *Administrative science quarterly*, pp. 78-108, 1986.

[16] P. M. Leonardi, and S. R. Barley, "Materiality and change: Challenges to building better theory about technology and organizing," *Information and Organization*, vol. 18, no. 3, pp. 159-176, 2008.

[17] S. R. Barley, J. Freeman, and R. C. Hybels, "Strategic alliances in commercial biotechnology," *Networks and organizations: Structure, form, and action*, pp. 311-347, 1992.

Using Prediction Markets to Motivate Public Participation in Patent Examination

[Extended Abstract]

Lian Jian*
School of Information
University of Michigan
ljian@umich.edu

ABSTRACT

The United States Patent and Trademark Office (USPTO) is overburdened with a large volume of patent applications while having limited resources to conduct patent examinations. The patent examination process is too long and the quality of issued patents is questioned by the public. I propose to alleviate these problems by setting up prediction markets for each pending patent. In these prediction markets, traders buy and sell bets for the outcome of the patent examinations. These proposed prediction markets can create social value in two ways. First, they generate forecasts about the likelihood of the pending patents being granted. Before the USPTO completes the examination, decision makers in need of information about the outcome of the patent examination can use these forecasts to make strategic decisions about research and development plans, or investments in the technologies being patented. Second, our proposal creates explicit incentives for public participation in the patent examination process. The proposed prediction markets reward traders with insights into the pending patent, potentially motivating traders to independently perform prior art search — a central task in evaluating patentability. The USPTO can then collect these prior art for reference by giving small rewards to traders who submit relevant prior art.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Design

Keywords

Intellectual Property, Patent, Markets

*I thank Rahul Sami and Jeffrey MacKie-Mason for their comments and suggestions.

1. INTRODUCTION

The patent system was set up to foster innovation by granting the inventors exclusive rights to extract monopolistic profits from their own inventions for a limited length of time (Article 1, Section 8 of the United States Constitution). Fundamentally, the patent system was based on the premise that patents were truly "inventions" that deserve such privilege. If a patent which was not truly innovative was granted by the USPTO, the system would impose a dead weight loss of efficiency on society due to the unjustified monopoly. Over the recent decade, the USPTO's performance in patent examination has come under sharp criticism, especially the long delays in the examination process and the low quality of the patents granted.

Patent examination takes a long time, and the time it takes has increased in recent years. In the last fourteen years, the average pendency — the time in months from filing to either issuance or abandonment — has been above 18 months. In 2008, the average pendency across all technological fields reached 32.2 months. In the field of Software & Information Security, it has reached 42.4 months — more than three years.[1] The long pendency of patent examination exacerbates the patent hold-up problem, which occurs when a firm has invested in developing a technology before it discovers it infringes on another firm's patent. As patent examination has been largely a secret process in which only the examiners and the applications are involved, the longer the pendency, the more likely that a hold-up problem will occur. Increased pendency also leads to high damages to the firm being held-up.

There have been controversies over the quality of some high-profile patents. A well-known one was Amazon's one-click shopping cart patent (US Patent 5,960,411), which was granted by the USPTO in 1999[13]. One possible reason for the decline of patent quality is that the USPTO is inundated with patent applications and has limited resources. The USPTO receives about 1,000 applications every working day [7, Ch. 5]. Each patent application receives about 20 hours of attention on average from its examiners [12, 2], sometimes as little as 8 hours [7, Ch. 5]. Further, patent examiners face particular challenges in identifying non-patent prior art, due to their lack of participation in the scientific community, thereby not being up-to-date on where the latest inventions are published [17].

[1]These data are published by the USPTO.

2. PUBLIC PARTICIPATION IN THE PATENT EXAMINATION PROCESS

Various reforms of the patent examination rules have been suggested [19, 18, 7], most of which require a long time to take effect. Entities other than the USPTO have initiated projects to bring the public into the patent examination process, achieving varying degrees of success. A few examples are BountyQuest (2000 $\sim$ 2003), Wikipatents (wikipatents.com) and Article One Partners [15].

My proposal is built on Peer-to-Patent (P2P), a pilot project launched by the USTPO in 2007, to harness the "wisdom of the crowd" to identify prior art for pending patent applications. For each patent application published on P2P, anybody can post prior art or vote for the most relevant prior art already posted. Four months after the listing of a pending patent on P2P, the USPTO collects the submitted prior art and considers them in their examination process. P2P achieved moderate success during its first year [1]. The first 27 patent examination decisions issued during the pilot phase showed use of P2P submitted prior art in nine rejections. Also, non-patent prior art was submitted to the patent office through the P2P site, compensating for the lack of expertise on the patent examiners' part on non-patent prior art search.

P2P has not gone without criticism [4, 5], particularly about the incentives to participate in this community. Currently P2P relies entirely on members' altruism to voluntarily contribute prior art, which is not a robust type of incentive. Especially for experts in specific technological fields, the opportunity cost of time can be high.

3. CREATING PREDICTION MARKETS FOR PATENT APPLICATIONS

I propose to implement a prediction market for each pending patent application. Such markets will reduce the long pendency problem by providing early assessment of the likelihood of issuance of the patents. More importantly, they will improve on the P2P project by creating explicit incentives to participate in the prior art search.

In a prediction market participants trade securities whose values depend on the outcome of future events. A prediction market for a patent facilitates trades on the security based on the USPTO's action on that patent.Prices in such markets fully aggregates all individual traders' private prediction [14], and reflects what the market as a whole "thinks" the probability of the patent being issued is.

Prediction markets' advantages in aggregating multiple individuals' private predictions have been demonstrated in a large number of markets. The Iowa Electronic Markets (IEM) consistently outperforms opinion polls in predicting the two-party vote shares of U.S. presidential elections [3]. In corporate settings, prediction markets outperform traditional forecasting methods such as face-to-face meetings [6] and surveys [8].

3.1 The thin market problem

Thin markets are markets in which only a small number of buyers or sellers are willing to transact at any given time. As a result, the market price may not reflect the true relation between supply and demand. My proposed prediction markets are likely thin markets, due to which the accuracy of their predictions might suffer. First of all, a large number of patent applications are filed to the USPTO each year [7]. There may be a large number of markets on the same site simultaneously. It is likely that each market receives a small number of trades. Second, most patents are in specialized fields, in which only a small number of experts have sufficient insights to participate in trading. Third, knowing there might be expert traders in the market, non-expert traders may hesitate to participate for fear of loss.

To avoid the thin market problem, I propose to implement the prediction markets using the market scoring rule (MSR) format, proposed by [9]. MSR based prediction markets solve the thin market problem by having a market maker — an automated trader who is ready to trade with anyone at any time. Even if there is only one interested trader, she can still trade with the market maker, and her private information can thus be elicited. This property is desirable in specialized markets in which only a small number of experts participate, such as our patent markets.

3.2 Submitting prior art

To encourage traders to submit prior art that they have discovered, I propose to augment the prediction market with a channel which allows submission of prior art. If someone already has found some prior art, it costs her very little to share it. The main incentive for conducting a prior art search comes from the potential to profit in the prediction market. Individuals can then be offered a small lump sum of money for sharing the prior art they have already found, if their contribution is cited by the USPTO.

This lump sum monetary reward may not even be necessary. Individuals can benefit from submitting the prior art they have discovered. Presumably, if an individual has discovered a useful piece of prior art, she would be trading toward the direction that the patent will be invalidated. If she submits the prior art she discovered, there is a higher chance that the USPTO will invalidate the patent, hence increasing the chance that she will profit in the prediction market.

4. IMPLEMENTATION ISSUES

A few issues remain to be considered before my proposed prediction markets can be implemented.

- **Manipulation**. Competitors of a patent applicant or the inventors themselves may try to manipulate the market price to influence the final patent issuance outcome or to misguide each others' decisions on research and development activities. Both theory and empirical evidence have show than attempts to manipulate the market would only hurt the accuracy of the market predictions temporarily, because the presence of manipulators creates opportunity for legitimate traders to profit [10, 11].
- **Disclosure**. Some may worry that should a patent fail to be granted, disclosure in a prediction market would have given competitors opportunity to steal the invention. This worry is unfounded [16]. Currently, patent

applications are published after 18 months, granted or not.[2] Further, patent protection applies retrospectively to the date of invention. Thus whoever tries to steal the technology in the review process runs the risk of being sued for infringement should the patent be granted.

5. SUMMARY

In this paper, I propose to build a prediction market for each pending patent application, to alleviate both the pendency and the quality problem of the patent examination process. Prediction markets are markets in which traders buy and sell bets for the outcomes of future events. In our case, these future events are the issuance or abandonment of the patent application. These prediction markets generate an aggregated prediction for the likelihood of each pending patent being granted, before the USPTO makes a decision. It can reduce the occurrence of hold-up problems and can also help in incentivizing the public to participate in the prior art search, thereby increasing the quality of issued patents.

6. REFERENCES

[1] N. Allen, J. Merante, Y. Tham, J. Ingham, B. S. Noveck, M. Webbink, B. Johnson, W. Stock, and C. Wong. Peer to patent first anniversary report, 2008.

[2] J. R. Allison and M. A. Lemley. The growing complexity of the united states patent system. *Boston University Law Review*, 82:77, Jan 2002.

[3] J. E. Berg, F. D. Nelson, and T. A. Rietz. Prediction market accuracy in the long run. *International Journal of Forecasting*, 24(2):285 – 300, 2008.

[4] P. L. Blog. Peer-to-patent expected to launch in April 2007, Feb 2007. Retrieved on Oct 20, 2008.

[5] P. L. Blog. Peer-to-patent: Live, June 2007. Retrieved on Oct 20, 2008.

[6] K.-Y. Chen and C. R. Plott. Information aggregation mechanisms: concept, design, and implementation for a sales forecasting problem, 2002. Caltech Social Science Working paper #1131.

[7] F. T. Commission. To promote innovation: The proper balance of competition and patent law and policy, 2003. Ch. 5, at 4.

[8] B. Cowgill, J. Wolfers, and E. Zitzewitz. Using prediction markets to track information flow: Evidence from google, 2008.

[9] R. Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119, Jan 2003.

[10] R. Hanson. Foul play in information markets. In R. W. Hahn and P. C. Tetlock, editors, *Information Markets: A New Way of Making Decisions*. AEI-Brookings Joint Center for Regulatory Studies, Washington, D. C., 2006.

[11] R. Hanson, R. Oprea, and D. Porter. Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior and Organization*, Jan 2006.

[12] J. L. King. Patent examination procedures and patent quality. In *Patents in the Knowledge-Based Economy*. National Academies Press, 2003.

[2]The Administrative Procedure Act ("APA"), 5 U.S.C. §553.

[13] E. Mills. Amazon 1-click patent rejected. *CNET News*, 2007. Retrieved on Oct 20, 2008.

[14] J. F. Muth. Rational expectations and the theory of price movements. *Econometrica*, 29:315–335, 1961.

[15] G. News.tv. Online startup aims to improve patent quality, Nov 17, 2008.

[16] B. S. Noveck. Peer to patent: Collective intelligence and intellectual property reform. *Harvard Journal of Law and Technology*, Jan 2006.

[17] B. N. Sampat. Examining patent examination: An analysis of examiner and applicant generated prior art. *stiy.com*, Jan 2004.

[18] C. Shapiro. Patent reform: Aligning reward and contribution. *NBER Working Paper*, Jan 2007.

[19] J. R. Thomas and W. H. Schacht. Patent reform in the 110th congress innovation issues. *ieeeusa.com*, Jan 2007.

The Imagined User of "Universal" Information Access Efforts: Ingrained Assumptions in Early American Public Libraries and Large-Scale Digitization Initiatives

Elisabeth A. Jones
The Information School
University of Washington
Seattle, WA 98195
eaj6@uw.edu

1. INTRODUCTION

Five years ago, two ambitious book-scanning initiatives – Google Book Search and the Open Content Alliance – were launched, both claiming the eventual goal of digitizing every book in the world, for the use of every person in the world. The initiatives have followed different paths: one private, one public; one centralized, one dispersed; one scanning everything right away, one starting with the public domain. Both, however, have been lauded for their groundbreaking potential to increase access to information worldwide.

Still, the basic impetus that underlies these initiatives is far from novel. In fact, the central motivation of such large-scale digitization initiatives (LSDIs) – to provide wide-ranging information access to as many people as possible – has strong historical precedents, especially in the early history of the American public library. Specifically, like LSDIs, early free public libraries reflected a top-down, supply-side approach to information access, and incorporated a high degree of private patronage at their initiation. The history of the American public library can thus illuminate many of the positive outcomes that can result from large-scale information initiatives; however, it also reveals some of the perils they might encounter.

In this poster, I will begin to explore one facet of the comparison between LSDIs and early free public libraries: that is, the sense in which each is constructed around a particular vision of "the imagined user," and how the inscription of that imaginary in each case has impacted – or might in the future impact – the claims to universality maintained by each.

2. IMAGINED USERS

2.1 The Early American Public Library

Tax-supported American public libraries were intended as broadly public institutions from their very beginnings in the mid-nineteenth century [1-4]. Still, the individual motivations that guided their structural and political design were not nearly so broad. In fact, the public library movement *per se* was largely built upon the assumptions and motives of a small set of wealthy, powerful, and often paternalistic white Anglo-Saxon protestant men. These early public library leaders shared many common assumptions; among them, two dueling visions of library users, as either (a) genteel, self-improving, perfectible and aspiring members of the middle class, or (b) more frightening specters, defined by their economic, ethnic, and intellectual "otherness."

2.1.1 The Genteel, Aspirational User

One vision of the library user espoused by early public library leaders was fairly optimistic: they assumed that library patrons would be well-mannered, interested principally in improving themselves, and if not already middle class, then at least aspiring to be so. To put it differently, the leadership imagined that the library-using populace would be like themselves, only less so; that they would aspire to be more like the wealthy industrialists and land barons who funded much early library development. Andrew Carnegie, for example, suggested that libraries would "stimulate the best and most aspiring poor of the community to further efforts for their own improvement" [5]. The masses would use libraries to learn, and thus mold themselves to conform with, if not the library leaders themselves, then at least the genteel middle class.

Yet, though they were imagined as *aspiring* members of the middle class, library users were expected to behave according to that class's genteel standards well before they had achieved those aspirations. This expectation, predictably, fostered alienation. As Garrison notes, by the turn of the century, laborers had the sense that, "the public library actually had been reserved, albeit 'unconsciously,' for members of the educated middle class – 'those who need it least and use it little'" [6].

2.1.2 The User as "Other"

The second early vision of the library user differs strikingly from the one above. Instead of the paternalistic optimism exhibited in the expectation of the genteel user, the vision of the user as "other" emerges from the more condescending presumption that library users would be everything elite library advocates were *not* – working-class, foreign-born, ill-educated, ill-mannered, and frankly, a little scary. This negative face of library leadership's paternalist tendencies thus cast users as uncivilized beings unable to advance themselves without the kindly assistance of their social betters; the users became not only "other," but also *lesser*.

The urge to educate, and thus to civilize, the "illiterate blacks and foreign born" was present among library leadership from the outset [2], and indeed, several historians have noted the ethnic chauvinism of various library founders. For example, Harris cites BPL Trustee George Ticknor's assertion that recent immigrants "at no time, consisted of persons who, in general, were fitted to understand our free institutions or to be intrusted with the political power given by universal suffrage" [3]. And

Garrison adds that an early president of the ALA, Charles Cutter, appeared to divide users into two types, "the fit and the unfit, the readers of The Nation and the hordes in the factories and tenements," and that he and other library leaders sought to use public libraries to preserve the socioeconomic status quo [6].

2.2 Large-Scale Digitization Initiatives

The imagined user of LSDIs remains far less clear than that of the early public library, mainly because of their newness, but also partially because of the private or semi-private initiatives' lack of transparency relative to taxpayer-supported public libraries. Still, the evidence that is available does begin to indicate a few of the imaginaries and assumptions at work in their design. To take one of the clearest examples, LSDI users, like previous digital library users, seem to have been imagined centrally as education-seekers, as opposed to entertainment- or social-interaction-seekers. In Sergey Brin's expressed desire to provide the "highest quality knowledge" [7] or Brewster Kahle's commitment to "living up to the dream of the Library of Alexandria and then taking it a step further" [8] one hears echoes of public library leaders' calls to educate and uplift the masses by providing "a better class of books than the ephemeral literature of the day" [9]. And with those echoes, one wonders whether there might also come a parallel condescension to those masses, or at least a parallel paternalism.

LSDIs also reveal their assumed user in other ways: through the languages in which the interfaces are offered, through the degree of technological expertise required to locate desired information, through the epistemological lines drawn by their classification systems, and even by the usage of the very Western library metaphor for their design. Each of these reveals a facet of the user profile envisioned by those shaping LSDI design, and each presents intriguing directions for future research.

3. IMPLICATIONS

The imagined aspirational user of the early American public library, once inscribed in the policies and architecture of the institution, recursively impacted the actual use of that institution. The imposing structures and genteel social norms of early libraries, reflective of the entrenched social hierarchies of the time, repelled the rhetorical target of the public library movement, the so-called "working man." In fact, it took several decades for the public library to divorce itself from the structural biases entrenched at the very beginning, by introducing open shelving and increased user input into collection development, as well as other more welcoming amenities.

Today's large-scale digitization initiatives may similarly be targeted at a more restricted audience than their rhetoric implies. By principally targeting an education-oriented audience, LSDIs risk blinding themselves to other valuable uses of information, such as social interaction, community building, and entertainment. Additionally, through their choices regarding language, technological accessibility, and design metaphors, LSDIs narrow the profile of their imagined user along each of those lines.

To the extent that designers of broad-scale information access initiatives – whether digital or analog – create systems with the potential to radically shift information practices worldwide, they have an obligation to consider how they might make these systems maximally inclusive of both diverse uses and diverse people. Public libraries are still not perfect in this regard, but they have made great progress – and their 150-year history has much to tell us about the possibilities and perils for newer efforts like LSDIs. This paper forms a starting point for research into these parallels and their implications.

4. REFERENCES

[1] Lee, R. 1971. The People's University—the Educational Objective of the Public Library. In Harris, M. H. Ed., Reader in American Library History. Microcard Editions, 117-124.

[2] Ditzion, S. 1947. Arsenals of a Democratic Culture. ALA.

[3] Harris, M. H. 1975. The Role of the Public Library in American Life: A Speculative Essay. University of Illinois Graduate School of Library Science.

[4] Shera, J. H. 1949. Foundations of the Public Library. University of Chicago Press.

[5] Carnegie, A. 1889. The Best Fields for Philanthropy. The North American Review, 149, 397, 682-698.

[6] Garrison, D. 1979. Apostles of Culture: the Public Librarian and American Society, 1876-1920. Macmillan.

[7] Toobin, J. 2007. Google's Moon Shot: The Quest for the Universal Library. The New Yorker (February 5).

[8] Kahle, B. (2005). Announcing the Open Content Alliance. Yahoo! Search Blog (October 2).

[9] Wadlin, H. D. (1911). The Public Library of the City of Boston: A History. Boston: Trustees of the Public Library.

Developing a Usability Measurement Instrument in Academic Digital Libraries

Soohyung Joo
School of Information Studies
University of Wisconsin-Milwaukee,
P.O. Box 413, Milwaukee, WI
+1-414-793-1748
sjoo@uwm.edu

Keywords
Usability, Digital libraries, Measurement instrument

1. INTRODUCTION
This poster introduces an ongoing research to develop a usability measurement instrument in the context of academic digital library. In the field of information science, there have been many research related to the evaluation of usability in different information environments. While most usability evaluation studies employed either inspection methods or user experiments, less research applied user survey methods. However, a user survey method could complement those two predominant methods, inspection and experimentation, in terms of involving large samples. To implement user survey methods, it is prerequisite to develop a reliable and valid measurement instrument. This study attempts to develop a measurement instrument for usability specific to academic digital library settings. In this study, the academic digital library refers to an augmentation of a traditional academic library, which includes electronic subscription, access to online database, self-digitized collections, and virtual references comprehensively.

2. MEASUREMENT FRAMEWORK
In this study, the International Standards Organization's (ISO) 9241-11 standard [7], which is one of most widely cited model in usability studies, was employed to build a theoretical ground. According to ISO 9241-11, usability is defined as '*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use* (p. 2) [7].' As the definition shows, the usability consists of three components – effectiveness, efficiency, and satisfaction. Effectiveness refers to the completeness at which users achieve specified goals; efficiency refers to the resources used in completing a task; and satisfaction refers to positive attitudes toward using the system [7]. However, the satisfaction is dependent on effectiveness and efficiency because users who work with highly effective or efficient information systems are likely to perceive more satisfaction. Frokjaer et al. [5] addressed the correlation between effectiveness, efficiency, and satisfaction. Thus, in order to come up a parsimonious instrument, only two subscales, effectiveness and efficiency, were adopted from the ISO 9241 standard because the satisfaction subscale could overlap with other subscales.

Instead, learnability was identified as an additional subscale. There are several research that identified learnability as one of attributes of usability [2, 6, 9, 10]. In this study, the definition by Nielson [9] will be used, which addresses how easy it is for casual users to learn a system.

With the construct of usability and three associated subscales, a path diagram was drawn (Figure 1). As seen in Figure 1, the usability measurement instrument consists of three subscales, effectiveness, efficiency, and learnability, and corresponding items.

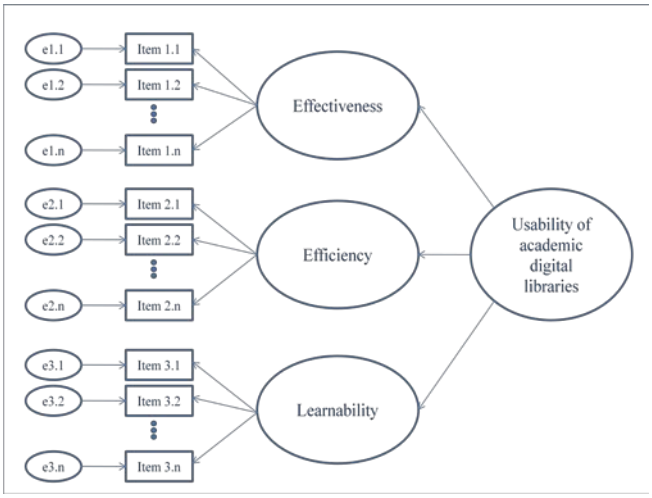


Figure 1: The path diagram of usability measurement instrument

3. THE INITIAL LIST OF ITEMS
On the basis of the framework identified above, an initial list of measurement items, which consists of three subscales and nineteen items, was generated. Table 1 presents the initial list of measurement items.

Table 1: The initial list of items

Subscale	Item
Effectiveness	I can usually complete a search task using this digital library.
	I am successful in general in finding academic resource(s) using this digital library.
	Overall, this digital library is useful in helping me find information.
	I usually achieve what I want using this digital library.
	The resources I obtain from this digital library are useful.
	I can trust the information obtained from this digital library.
	This digital library covers sufficient topics that I try to explore.
Efficiency	It is easy to find the academic resources that I want on this digital library.
	This digital library is easy to use in general.
	I can complete a resource finding task quickly using this digital library.
	This digital library is well designed to find what I want.
	It is easy to perform searches on this digital library.
	I get the results of searches quickly when using this digital library.
Learnability	It was easy to learn to use this digital library.
	The terminologies used on this digital library are easily understandable.
	This digital library offers easy-to-understand menus.
	This digital library has appropriate help functions.
	This digital library provides well-organized help information for new users.
	It does not take a great deal of effort for new users to become proficient with this digital library website.

4. RELIABILITY EVALUATION PLAN

This research plans to evaluate three kinds of different reliability: (1) internal consistency reliability; (2) item reliability; and (3) construct reliability. First, Cronbach's alpha coefficients will be calculated to examine the internal consistency reliability of the initially identified items. Second, for the item reliability evaluation, a structural equation modeling (SEM) will be applied to get the multiple squared correlation between each item and its associated subscale. Third, the construct reliability, which refers to the degree to which the measurement of the set of latent items of a construct is consistent [8], will be also examined on the basis of SEM outputs.

5. VALIDITY EVALUATION PLAN

5.1 Content Validity Examination

Content validity indicates the extent to which the items comprehensively represent the identified construct [3]. Generally, the most widely used technique for content validity is expert judgment. Experts can examine whether the items properly account for a construct as intended. In this study, at least, three scholars in the information system evaluation area will review the initially identified items. Following criteria are supposed to be used to examine the content validity:

• How properly each item accounts for a subscale as intended?

• How clearly are the construct and its related subscales defined?

• How clear is the statement of each item?

5.2 Construct Validity Examination

Construct validity indicates the extent to which an item accurately measures the construct of interest [11]. In this project, three kinds of analyses are planned to be conducted to evaluate the construct validity of the initially identified measurement items.

First, a correlation analysis between all items will be conducted. In general, a well-structured measurement instrument has relatively higher correlation coefficients between the items that belong to the same subscale, whereas it has comparatively lower correlation coefficients between the items that belong to different subscales. Second, an explorative factor analysis will be conducted to identify an optimal set of measurement items. Third, a confirmatory factor analysis (CFA) will be implemented using SEM. Using CFA analysis, both of the convergent validity and the discriminant validity will be evaluated.

6. SUMMARY

As of December 2009, the present research has generated the initial pool of measurement items specific to the academic digital libraries. The initially identified items are subjected to a comprehensive examination of reliability and validity empirically. Thus, a survey is planned to gather the data from the actual users of academic digital libraries. Based on stratified sampling method, a sample will include more than 200 users involving faculties, students, and staffs. After verifying the reliability and validity, a final set of measurement items is expected to be suggested from this project.

7. REFERENCES

[1] N. Bevan, and M. Macleod. Usability Measurement in Context. Behavior & Information Technology, 13:132-145, 1994.

[2] T. Brink, D. Gergle, and S.D. Wood. Designing Web sites that work: Usability for the Web. Morgan Kaufmann, San Francisco, 2002.

[3] M.R. Carrier, A.T. Dalessio, and S.H. Brown. Correspondence between estimates of content and criterion-related validity values. Personnel Psychology, 43:85-100, 1990.

[4] T. Cook, and D. Campbell. Quasi-experimentation: Design and analysis issues in field settings. Houghton Mifflin, Boston, 1979.

[5] E. Frokjaer, M. Hertzum, and K. Hornbaek. Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? CHI 2000 Conference on Human Factors in Computing System, The Hague, The Netherlands, April 1-6, 2000.

[6] K. Guenther. Assessing Web site usability. Online, 27: 65-68, 2003.

[7] International Standard Organization (ISO). 'ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs)', Part 11: Guidance on Usability Specification and Measures. Technical report, 1997.

[8] C. Lu, K. Lai, and T.C.E. Cheng. Application of structural equation modeling to evaluate the intention of shippers to use Internet services in liner shipping. European Journal of Operational Research, 180(2): 845-867, 2007.

[9] J. Nielson. Usability Engineering. Academic Press, Cambridge,1993.

[10] B. Shackel. Usability – Context, framework, definition, design and evaluation. In Shackel, B & Richardson, S. (Eds.), Human Factors for Informatics Usability, Cambridge University Press, Cambridge, UK, pages 21-37, 1991.

[11] M.R. Wade. and S. Nevo. Development and validation of a perceptual instrument to measure e-commerce Performance. International Journal of Electronic Commerce, 10(2):123-147, 2005.

Effects of Ease of Use, Effectiveness, and Use Frequency on User Satisfaction in Academic Library Website Uses

Soohyung Joo
School of Information Studies
University of Wisconsin-Milwaukee,
P.O. Box 413, Milwaukee, WI
+1-414-793-1748
sjoo@uwm.edu

1. INTRODUCTION

For undergraduate and graduate students, library website services such as an online library catalog, electronic resources, and online reference services are considered imperative in achieving their academic goals. To satisfy student users, it is important to develop a user-friendly website based on precise understandings of factors associated with user satisfaction. There are various factors affecting satisfaction such as availability, resource quality, usability, and service quality, and many studies have investigated causal relationships between various factors and user satisfaction in the context of academic library web-based services. This research investigated the effects of three variables, ease of use, effectiveness, and use frequency, on satisfaction in using a university library website.

2. RESEARCH QUESTION

The research question was established as: how use frequency, ease of use, and effectiveness affect users' satisfaction in uses of university library websites?

3. METHODOLOGY

3.1 Sampling

The participants of this research were students of Yonsei University (www.yonsei.ac.kr) in Korea. The number of valid respondents was 191, which consist of 119 undergraduate students and 72 graduate students.

3.2 Operationalization of Concepts

This study identified four concepts: (1)frequency of use, (2)ease of use, (3)effectiveness, and (4)satisfaction. Each concept was operationalized as following:

• *Use frequency* refers to how frequently a user uses a library website.

• *Ease of use* refers to the extent how easy a user perceives a library website when using it.

• *Effectiveness* refers to how successful is when a user performs his/her specific information seeking tasks.

• *Satisfaction*: Freedom from discomfort, and positive attitudes towards the use of the website [1].

3.3 Measurement

In this research, three independent variables and one dependent variable were identified:

• X0 (categorical variable; coded as E1 and E2 vectors) – Use frequency

• X1 (continuous variable) – Ease of use

• X2 (continuous variable) – Effectiveness

• Y (continuous variable) – Satisfaction

The variable of "use frequency" was a categorical variable that indicates how frequently a participant uses a university library website. This categorical variable consists of three different categories: (1)more than once a week; (2)more than once a month (but less than once a week); and (3)less than once a month or never use. For this poster, each category is named as "group 1" for "more than once a week", "group 2" for "more than once a month" and "group 3" for "less than once a month or never use". The other three variables – "ease of use," "effectiveness," and "satisfaction" – were continuous variables, and the respondents were asked to rate their perceptions using seven-point Likert scale. For simplicity, each variable is presented using a code – X0 (E1 and E2), X1, X2, and Y – respectively

3.4 Strategy of Analysis

To answer the research question, this study adopted a multiple regression with three independent variables involving one categorical variable and two continuous variables. This regression model also examined whether the significant interaction effects occurred between X0 and X1, X2. The Figure 1 shows the research design of this study. For the coding of categorical variable that has three categories, an effect coding method was employed.

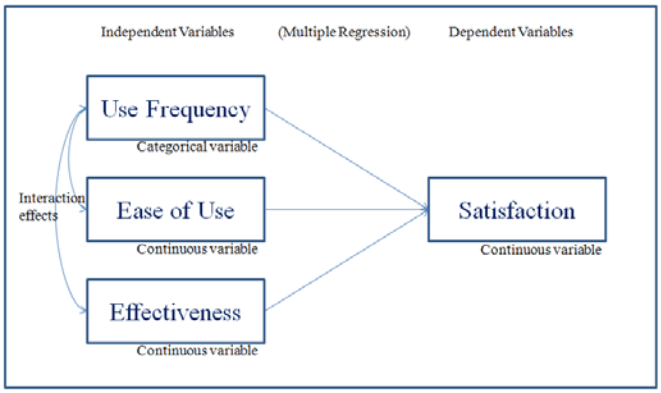


Figure 1: Research design

4. RESULTS

A descriptive statistics of each variable was investigated. 57.9% of participants answered they used library website more than once a week. 23.6% of respondents used more than once a month, but less than once a week, and 18.5% answered less than once a month. For the continuous variables, the mean values of X1, X2, and Y were 3.91, 4.26, and 5.30, respectively.

To investigate how X0(use frequency), X1(ease of use), X2(effectiveness) affect Y(satisfaction), a multiple regression was conducted. For the X0, an effect coding using two vectors (E1 and E2) was applied. Then, the independent variables were entered into the analysis using three different blocks to see the R-square change to find the most appropriate regression model. Three different models of regression were identified(Table 1): "model 1" included only E1 and E2 (vectors of X0); "model 2" involved E1, E2, X1, and X2; and "model 3" included E2, E1, X1, X2, and the interaction effects between E1 and E2 and X1 and X2. Table 1 shows the R-square of each model and its change. "Use frequency (E1 & E2)," "ease of use (X1)," "effectiveness (X2)," and "their interaction effects" accounted for approximately 22.6% of the variance of "satisfaction (Y)." The F change from "model 2" that included main effects of X0, X1 and X2 to "model 3" that included three main effects and the interaction effects of E1 and E2 and X1 and X2 turned out to be 1.311, but it was not significant at the 0.05 alpha level.

Table 1. R-squares and their changes in regression models

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.238[a]	.057	.047	1.303	.057	5.667	2	188	.004
2	.451[b]	.203	.186	1.204	.147	17.108	2	186	.000
3	.475[c]	.226	.192	1.200	.022	1.311	4	182	.268

a. Predictors: (Constant), E2, E1

b. Predictors: (Constant), E2, E1, X1, X2

c. Predictors: (Constant), E2, E1, X1, X2, Interaction of E1*X2, Interaction of E1*X1, Interaction of E2*X2, Interaction of E2*X1

Since the interaction effects between E1 and E2 and X1 and X2 were not significant, a multiple regression with the independent variables of X0 (E1 & E2), X1 and X2 was conducted, excluding interaction terms (Table 2). The result shows that the satisfaction scores were regressed on use frequency, ease of use, and effectiveness. These predictors accounted for approximately 20% of the variance in satisfaction scores, which was significant, $F(4,186) = 11.873$, $p<.05$. Both the ease of use ($b = .153$, $p<.05$) and the effectiveness ($b=.373$, $p < .05$) demonstrated significant effects on the satisfaction scores.

Table 2 Coefficients of regression model

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2.994	.414		7.227	.000		
	E1	.307	.121	.176	2.539	.012	.893	1.120
	E2	-.111	.146	-.053	-.760	.448	.892	1.121
	X1	.153	.072	.139	2.117	.036	.998	1.002
	X2	.373	.068	.374	5.505	.000	.928	1.077

a. Dependent Variable: satisfaction

To analyze further, the separate regression equations were calculated for three different groups by use frequency. Since the coefficients of main effects of X1 and X2 were significant ($p<.05$), a regression equation for each group by use frequency was able to be obtained. The overall regression equation for three identified groups, not including the interaction term, was "$Y' = 2.994 + 0.307*(E1) + (-0.111)*(E2) + 0.153*(X1) + 0.373*(X2)$." From this, three different separate regression equations were obtained, which have common regression coefficients, but different intercepts:

- (For group1): $Y' = 3.301 + 0.153*(X1) + 0.373*(X2)$
- (For group2): $Y' = 2.883 + 0.153*(X1) + 0.373*(X2)$
- (For group3): $Y' = 2.798 + 0.153*(X1) + 0.373*(X2)$

These separate regression equations revealed that the participants who used the library web services more frequently would be likely to perceive higher satisfaction if the conditions of X1 and X2 are same. From this multiple regression, we conclude that use frequency, ease of use, and effectiveness would affect user satisfaction in using university library website, but there were no interaction effects between these three independent variables.

5. CONCLUSIONS

This study investigated how use frequency, ease of use and effectiveness affect the users' satisfaction in the uses of university library web services. The result of regression analysis shows that use frequency, ease of use, and effectiveness would affect the satisfaction, but there was no significant interaction effect. The main effect of effectiveness was more influential on the satisfaction than the ease of use.

These findings yield practical implications for web-service designs for academic libraries. First, effectiveness is more closely related to user satisfaction than ease of use, so when developing a web-based service in academic library, the effectiveness could be more emphasized than ease of use to improve user satisfaction. Also, users who use the library website less frequently are likely to be less satisfied. Thus, to enhance the satisfactory levels of users with less frequent uses, it will be required to understand specific needs of them.

However, this research also has its limitations. The R-square value could be enlarged to explain more proportion of the satisfaction. To do this, it is needed to involve more factors affecting the satisfaction. The three variables investigated in this research are not sufficient to explain sufficient proportions of the satisfaction. Further research can be designed to include more independent variables such as availability, service quality, and others.

6. REFERENCE

[1] International Standard Organization (ISO). 'ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs)', Part 11: Guidance on Usability Specification and Measures. Technical report, 1997.

A Web Link Structure of the American Library & Information Science Field: A Pilot Study

Soohyung Joo
School of Information Studies
University of Wisconsin-Milwaukee,
P.O. Box 413, Milwaukee, WI
+1-414-793-1748
sjoo@uwm.edu

Keywords
Web link analysis, Social network analysis, Webometrics, Library and Information Science Field

1. INTRODUCTION
Web link analysis has become one of the methods to delineate a communication structure among different scholars or organizations with the advent of the Web. This study represents a pilot investigation of the Web space in the Library and Information Science (LIS) field using social network analysis techniques. By relying on collected in- and out-link data using a hyperlink crawler and analyzing the directed hyperlink data between Web sites in the LIS field using different techniques, this study tried to explore the structure of the Web space in the field of LIS. Also, this study investigated which Web sites play a central role in communication in the Web space, along with the dependency between Web sites. As an exploratory study, the entire breadth of LIS-related Web sites was not covered.

2. RESEARCH QUESTIONS
This study attempts to address two research questions in relation to Web space analysis of the LIS field:

1) Which sites will be located in the central region of the LIS Web space, as measured by link analysis?

2) What types of dependency patterns emerge among the Web sites?

3. METHODOLOGY

3.1 Data Collection
Twenty-four Web sites were selected, which include relevant scholarly and professional organizations, government institutions, and American LIS schools. For scholarly and professional organizations, representative organizations were chosen such as the American Library Association (ALA), the American Society for Information Science & Technology (ASIS&T), and the International Federation of Library Associations and Institutions (IFLA). Fourteen LIS schools accredited by ALA were also selected. Table 1 shows the 24 research objects in this study, and the ID (code) used in an italic font to represent the Web site of the institution.

Table 1: The list of 24 research objects

Type	Institution	ID (Code)
Organization	American Library Association	*ala*
	The International Federation of Library Associations and Institutions	*ifla*
	Library of Congress	*loc*
	Online Computer Library Center	*oclc*
	Special Libraries Association	*sla*
	Institute of Museum and Library Service	*imls*
	The American Association of Law Libraries	*aall*
	The National Archives and Records Administration	*archives*
	The American Society for Information Science & Technology	*asis&t*
	The Society of American Archivists	*saa*
School	School of Library and Information Science, The Catholic University of America	*slis.cua*
	School of Library and Information Science, Indiana University Bloomington	*slis. indiana*
	School of Library and Information Science, Kent State University	*slis.kent*
	School of Information, University of Michigan Ann Arbor	*si.umich*
	School of Information Sciences, University of Pittsburgh	*is.pitt*
	School of Communication, Information and Library Studies, Rutgers	*scils. rutgers*
	Graduate School of Library and Information Science, Simmons University	*lis. simmons*
	School of Information, The University of Texas Austin	*is. utexas*
	Department of Information Studies, UCLA	*is.ucla*
	Graduate School of Library and Information Science, UIUC	*lis.uiuc*
	School of Information and Library Science, UNC	*sils.unc*
	School of Information Sciences, University of Tennessee Knoxville	*sis.utk*
	College of Information Science, Florida State University	*ci,fsu*
	Information School, University of Washington	*is.washington*

In- and out-link data for the 24 Web sites were collected using a Web crawler that was designed to collect the hyperlinks to external domains. The data collection was conducted in February 2009. The crawling depth was set to the fourth level in the study. Even though the fourth level of depth did not necessarily reflect the whole structure of Web site, it would be acceptable for an initial investigation of this topic.

3.2 Analysis Techniques

For this pilot study, two approaches were identified.

1) *In-link and Out-link Visualization*: First, to understand the overall Web space structure of the LIS field, the in- and out-link data were visualized using Netdraw software (http://www.analytictech.com/Netdraw/netdraw.htm), which was developed for social network studies.

2) *Degree Centrality Test*: For more detailed investigation of the characteristics of each node, a degree centrality test was included. Centrality in network analysis can be used to quantify an individual node's prominence, influence or dependency using in-degree and out-degree measures. In-degree refers to the number of in-links, whereas out-degree refers to the number of out-links. In network analysis, in-degree represents how prestigious a node is because it indicates how many links a node receives from other nodes[1].

4. RESULTS

An in- and out-link matrix was creating using the external link crawler. The total number of hyperlinks (in-links and out-links) among the twenty four nodes up to the fourth level of each site was 7,898. Table 2 summarizes the descriptive statistics of the asymmetric in- and out-link matrix. The average number of hyperlinks between two nodes was 14.31 with a standard deviation of 41.39. The *ala* node had the largest number of in-links and out-links, totaling 2,156 and 1,147 links, respectively. This reveals that the ALA site is the most frequently linked with others in the LIS Web space. The second largest node was *loc*, with 1,247 in-links and 1,122 out-links. Third was *asis&t* with a total of 1,096 in- and out links. The matrix revealed that the average number of in- and out-links was larger in organizational nodes rather than in school nodes. The average number of links of each organizational node was 474.13, whereas the average links of each school node was 212.65. From this we can deduce that the organizational nodes have more in-links (6,683) than out-links (4,222), while the school nodes have more out-links (3,676) than in-links (1,215).

Table 2: Descriptive statistics of in- and out-link asymmetric matrix data

		(A) in-links from other 23 nodes		(B) out-links to other 23 nodes		(C) Total (A+B)	
		Counts of (A)	Ratio	Counts of (B)	Ratio	Counts of (A+B)	Ratio
.org/.gov	*ala*	2156	27.3%	1147	14.5%	3303	20.9%
	asis&t	405	5.1%	691	8.7%	1096	6.9%
	ifla	412	5.2%	213	2.7%	625	4.0%
	loc	1247	15.8%	1122	14.2%	2369	15.0%
	oclc	820	10.4%	37	0.5%	857	5.4%
	sla	540	6.8%	190	2.4%	730	4.6%
	imls	360	4.6%	44	0.6%	404	2.6%
	aall	150	1.9%	280	3.5%	430	2.7%
	archives	514	6.5%	122	1.5%	636	4.0%
	saa	79	1.0%	376	4.8%	455	2.9%
	Sub-total	6683	84.6%	4222	53.5%	10905	69.0%
.edu	*ci.fsu*	32	0.4%	259	3.3%	291	1.8%
	slis.cua	11	0.1%	205	2.6%	216	1.4%
	slis.indiana	93	1.2%	800	10.1%	893	5.7%
	slis.kent	118	1.5%	381	4.8%	499	3.2%
	si.umich	111	1.4%	63	0.8%	174	1.1%
	is.pitt	49	0.6%	144	1.8%	193	1.2%
	scils.rutgers	84	1.1%	137	1.7%	221	1.4%
	lis.simmons	12	0.2%	235	3.0%	247	1.6%
	is.utexas	106	1.3%	198	2.5%	304	1.9%
	is.ucla	87	1.1%	341	4.3%	428	2.7%
	lis.uiuc	140	1.8%	609	7.7%	749	4.7%
	sils.unc	294	3.7%	98	1.2%	392	2.5%
	sis.utk	13	0.2%	144	1.8%	157	1.0%
	is.washington	65	0.8%	62	0.8%	127	0.8%
	Sub-total	1215	15.4%	3676	46.5%	4891	31.0%
Total	Sum	7898	100.0%	7898	100.0%	15796	100.0%
	Mean	329.08	-	329.08	-	658.17	-
	SD	488.98	-	318.00	-	732.20	-

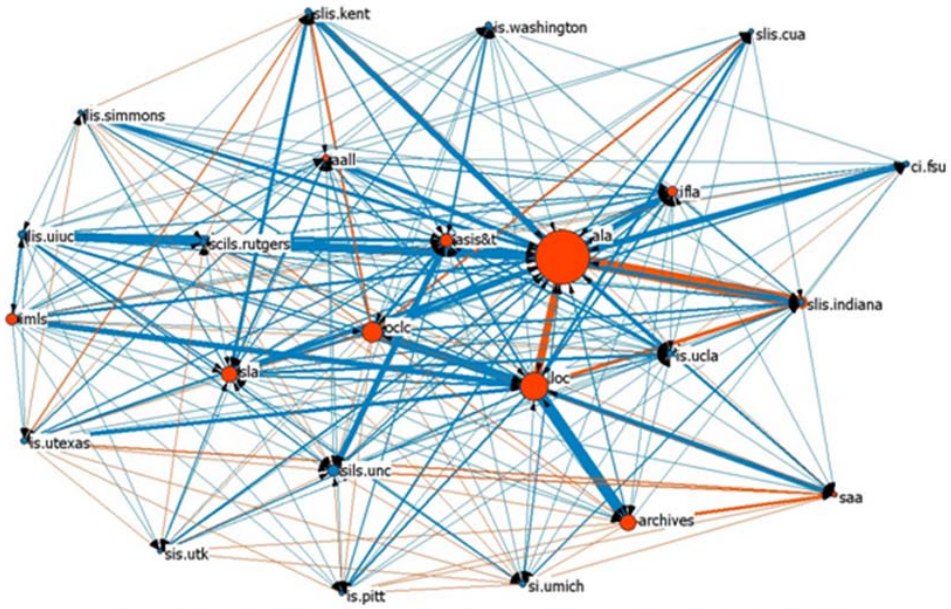


Figure 1: A linkage visualization of a Web structure of the LIS field in America

Figure 1 shows the visualization of the links among the twenty-four nodes. For the layout of nodes, the spring embedding algorithm, which provides an easily legible layout based on node repulsion concept, was applied with equal edge lengths. Since the spring embedding algorithm method considers distance as dissimilarity between nodes, we could interpret similarity among nodes from the output. In this figure, the red node indicates the organizational node and the blue node indicates the school node, and the size of node suggests the number of in-links. The arrow line indicates the link and its direction, and the width of the line reflects the frequency of links between two nodes.

The most notable finding was that most organizational nodes such as *ala*, *asis&t*, *oclc*, *loc* and *ifla* were located at the center of the space, whereas most school sites such as *slis.cua*, *ci.fsu*, *slis.kent*, *lis.simmons*, *sis.utk* and *is.pitt* were located at the periphery. A relatively distinct core/peripheral structure was apparent, where the core area consisted of organizational sites and the peripheral area consisted of school sites. Then the most prominent node was ala which had the largest size of node and the most frequent connections with other nodes. This suggests the ALA site serves as a core communicator in the LIS Web space. In addition, we can see the *ala* node sends and receives links to all other nodes, and the widths of those lines are relatively thicker. Furthermore, the four representative LIS related organizations, *ala*, *asis&t*, *oclc* and *loc*, located in close proximity to one another. Also, we can observe that *archives* and *saa*, which relate to the archives field, are located adjacently.

Table 3: Freeman's degree centrality measures

	(A)InDegree	(B)OutDegree	(C)NrmInDeg	(D)NrmOutDeg	(A) – (B)
ala	2156	1147	22.808	12.134	1009
loc	1247	1122	13.192	11.869	125
oclc	820	37	8.674	0.391	783
sla	540	190	5.712	2.01	350
archives	514	122	5.437	1.291	392
ifla	412	213	4.358	2.253	199
asis&t	405	691	4.284	7.31	-286
imls	360	44	3.808	0.465	316
sils.unc	294	98	3.11	1.037	196
aall	150	280	1.587	2.962	-130
lis.uiuc	140	609	1.481	6.442	-469
slis.kent	118	381	1.248	4.03	-263
si.umich	111	63	1.174	0.666	48
is.utexas	106	198	1.121	2.095	-92
slis.indiana	93	800	0.984	8.463	-707
is.ucla	87	341	0.92	3.607	-254
scils.rutgers	84	137	0.889	1.449	-53
saa	79	376	0.836	3.978	-297
is.washington	65	62	0.688	0.656	3
is.pitt	49	144	0.518	1.523	-95
ci.fsu	32	259	0.339	2.74	-227
sis.utk	13	144	0.138	1.523	-131
lis.simmons	12	235	0.127	2.486	-223
slis.cua	11	205	0.116	2.169	-194

To investigate the dependency patterns in the LIS Web space, the Freeman's centrality degree index values were calculated for each node. The higher in-degree centrality indicates a node gets higher attention by other nodes, whereas the higher out-degree centrality implies a node is more like to be dependent on other nodes in a Web space[1]. Table 3 shows the degree centrality in the LIS Web space. Regarding the in-degree centrality, the top eight nodes turned out to be the organizational sites. Similar to the outputs of the descriptive analyses, we see that the organizational sites receive higher numbers of in-links from other sites. Also, we find that most organizational nodes have higher in-degree centrality than out-degree centrality. Conversely, we find that most of the school sites have higher out-degree centrality than in-degree centrality. This pattern implies that the school sites are dependent

on other sites in the Web space. In particular, the *slis.indiana* node ranks third in out-degree centrality but fifteenth in the in-degree centrality, which indicates high dependency. From this degree centrality analysis, we are able to confirm the dependency propensity of school sites on organizational sites.

5. CONCLUSION

This study investigated the Web space of the LIS field using social network analysis techniques. The findings show that the organizational sites have more in- and out-links than school sites. The visualized network diagram of the LIS Web space imply most organizational sites are located near the center region, whereas most school sites are positioned in peripheral areas. Also the centrality analyses reveal that higher prominence of organizational sites than school sites and higher dependency propensity of school sites than organizational sites.

The findings of this study yield some insights in Web site evaluations, especially how to incorporate Web space structure in a specific field to evaluate the importance of Web sites in a specific field. How a specific Web site is situated in a Web space suggests a dynamic approach to evaluate the importance or influence of Web sites in relation to other sites. Also, the application of centrality measures such as degree, closeness and betweenness centrality utilized in social network analyses is useful for evaluating Web sites.

This study also has its limitations. First, even though the sample represents a selection of organizations and schools from the LIS field, the number of research objects could be enlarged to provide a more complete picture of LIS Web space. Second, the data did not represent the entirety of each site since the limitation of the crawling level. For more complete analysis, the depth of crawling could be deeper to reflect the full content of Web sites. Based on this pilot study, further research to implement Web space analysis with a larger dataset is anticipated.

6. REFERENCE

[1] S. Wasserman, and K, Faust. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994.

Meta Organizational Influences on Scientific IT Infrastructure Development

Kerk F. Kee
The University of Texas at Austin
1 University Station A1105
Austin, TX 78712-0115
512-487-1783
kerk.kee@gmail.com

ABSTRACT
By taking an organizational communication approach to examine the meta influences on scientific IT infrastructure development, a preliminary analysis reveals that existing practices in a field with the Internet and computer technologies, the agenda of the funding agency, and the competing theories and methodologies held by participating scientists and groups are three such meta organizational influences. Instead of presenting key findings in the form of statements, the student author instead raises meta questions to be asked as we develop and design large-scale scientific IT infrastructure in the early 21st century.

Categories and Subject Descriptors
J.4 [**Social and Behavioral Sciences**]: *Sociology*

General Terms
Management

Keywords
Cyberinfrastructure, IT infrastructure development, IT design, organizational communication, and meta influences

1. INTRODUCTION
Cyberinfrastructure (CI) refers to scientific IT infrastructure based on a collection of information, communication, computer technologies [1]. According to Stewart, "Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible" [10]. In the influential *Atkins Report*, Atkins and colleagues [1] further state that CI is "an effective and efficient platform for the empowerment of specific communities of researchers to innovate and eventually revolutionize what they do, how they do it, and who participates". Therefore, cyberinfrastructure represents a collection of machines and humans, as well as the social interactions and organizational practices surrounding the meshing of the two. CI development is inherently social and organizational. A conversation about IT infrastructure can include an examination of the meta organizational influences related to CI development.

2. LITERATURE REVIEW
One approach to examine the meta organizational influences around CI development is to employ the lens of organizational communication. An organizational communication approach treats communication as a way to explain the production of "social structures, psychological states, member categories, knowledge and so forth rather than … simply one phenomenon among these others in organizations" [3]. An organizational communication approach emphasizes the process of organizing through symbolic interaction [4]. In this paper, I attempt to identify a few sources of organizational influences on CI development by taking an organizational communication approach. In so doing, I highlight three organizational issues that communicate influences to CI development. This poster is based on three representative excerpts drawn from a data set of 65 interviews with domain scientists, computational technologists, supercomputer center administrators, social scientist and policy experts across 17 US states. I present the preliminary findings in terms of questions to be asked as we develop scientific IT infrastructure. I will present these questions in the following paragraphs.

3. FINDINGS
3.1 Existing Practices
The first question asks, "*What is a field's existing practices with the Internet and computer technologies?*" The Internet, computers, and a wide range of emerging information technologies shape today's organizational life [5, 7, 8]. Every scientific field has an existing set of practices with the Internet and computer technologies at work. This set of practices will affect how scientists in a field approach CI, and how CI development will impact their work. The more integrated the Internet and computer technologies are within the field, the more receptive the scientists will be to using CI. As an interview participant shares,

> [L]et's take bioinformatics [as an example]… The use of the Internet to do the science is dominant… [B]ioinformatics was born 10 years ago. So it grew up as the Internet was growing up. So biology almost started doing cyberinfrastructure without thinking… That field is richly cyber-enabled…. Science is evolutionary... If the previous step was on the Internet, the next step probably has to be on the Internet, by definition. So you're not able to not do cyberinfrastructure." (Professor of Informatics, Indiana).

As we examine scientific IT infrastructure development, we need to distinguish the different disciplines of science, and take into consideration a field's existing set of practices with the Internet and computer technologies. Scientific practices within a field are shaped by its history, and these practices can only change by slowly evolving over a long period of time. The design of CI should closely match existing practices, if CI is to be adopted to support a particular branch of science. Compatibility [6] with existing practices is key.

Furthermore, an effective design for one field may not be equally useful for another. Different fields developed their unique ways of doing science, and these organizational practices rooted in past successes are difficult to change. The CI development has to acknowledge the complexity and diversity in a wide range of disciplines and fields in science. If we need to build more CI to support different branches of science, funding is a key organizational issue to consider next

3.2 Funder's Agenda

The second question to consider asks, "*Who is funding the CI project, and which agenda is the project advancing?"* Although science is often assumed to be a neutral endeavor simply to improve human conditions in the society, the meta organizational influence associated with funding agencies behind the scene is not neutral nor value free. Funding agencies only fund projects that promote and advance their missions and agendas. If an agency is to fund a particular project, the money is given only to conduct science relevant to the agenda of the funding agency, and not the agenda of another. Below is what a policy expert reveals,

> "In some cases, in larger institutions, the problems are really magnified by the fact that faculty receives grants from NIH and other places, which did not encourage collaboration and joint usage of technology, but rather, waived them off and said – If you go out and get a Sun workstation on your desk, put a couple of Condors together, and then when the funding runs out, you're left with this big bill to run these machines… if it's [the funding is] going to be used for anything other than the research you did initially." (Policy Expert, Washington DC)

Funding is perhaps the most powerful driving force behind large-scale science in the US. A scientific IT infrastructure project is very expensive, and without funding from agencies such as NSF and NIH, no CI can be built. These funding agencies allocate resources to CI projects on a limited term basis. Once the allocated funding is used up, and if the project cannot secure continuing support, the operation comes to an end, including CI development for the project. In addition, the agenda of the funding agency influences CI development to prioritize activities in scientific research. What does not serve the agenda does not get built into the design.

Furthermore, funding is not neutral. By receiving funding from a particular agency, acceptance of the agency's agenda is implied. Therefore, while discussing scientific IT infrastructure development, it is important to keep in mind the political priorities communicated through funding to a particular project. If a CI project is to continue, the project has to continue advancing the agenda. CI development is not only closely aligned with the funding agency's agenda, it is also closely tied to the theoretical and methodological competitions within a field. This observation turns us to the third question.

3.3 Competing Theories or Methods

The third question asks, "*For which theory or method is the CI built?*" There are competing theories and methodologies within any disciplines and fields in science. A vibrant scientific community engages in a healthy debate about the different ideologies and approaches to doing science. However, when it comes to scientific IT infrastructure design and development, we inevitably encounter the competition among these different groups of scientists who hold different philosophies of science. As the last informant in this position paper points out,

> "We've been in disputes with people essentially having two different – not quite theories, but two methodologies to approach a problem. They would come to the cyberinfrastructure folks and say – We're glad to be on the project and of course you're going to include my methodology in the way the software works and exclude my competitor over there." (Supercomputer Center Administrator, Illinois)

The decisions made before and during the process in which CI is being developed to support science involve persuasions, arguments, or even conflicts between groups. The theory or methodology selected to guide CI design determines which theoretical and methodological camp gains ground in advancing its approach to science. Scientists compete to influence CI development in favor of their own orientation, and persuade computational technologists to write codes and build applications that will support their method. This is a process to indirectly weed out competing theories and methodologies in the field. CI design and development become a contested terrain among competing groups of scientists.

4. CONCLUSIONS

In this paper, I attempted to employ an organizational communication approach to highlight three sources of meta organizational influence that could affect scientific IT infrastructure design and development. Through preliminary analysis of selected excerpts from a larger interview data set, I presented three questions to consider while designing and developing CI: "*What is a field's existing practices with the Internet and computer technologies?*"; "*Who is funding the CI project, and which agenda is the project advancing?";* and "*For which theory or method is the CI built?*" These questions reveal that the existing practices of a field, funder's agenda, and

theoretical/methodological commitment of scientists can influence decisions that go behind CI design and development. In other words, the development of an infrastructure for organizing information, knowledge, and people cannot be free of influence by the cultural norms of a community, the agenda of the agencies funding the projects, and the sometimes incompatible ontologies/taxanomies within a knowledge domain.

This paper extends Star and Bowker's argument of infrastructure as an "installed base" [9] to consider a recursive relationship between organizational forces and infrastructure design. Star and Bowker contend a new technology "wrestles with the inertia of the installed base and inherits strengths and limitations from the base". This paper highlights the meta organizational influences that get built into the 'installed base' for future science and IT infrastructure development.

Moreover, this paper shows that scientific IT infrastructure design is political and complex. Bijker calls scholars to pursue "political questions" [2]. By employing an organizational communication lens and presenting the findings in a form of questions, this poster reveals the political-cultural relevance of meta organizational influences in IT infrastructure development.

A few implications can be drawn from the analysis. First, given the limited resources to build CI, the design is best to be flexible in order to adapt to a wide range of scientific fields. When there are discipline specific requirements, parts of CI can be built as extensions to cater to these needs. Second, CI projects may benefit from staying with one primary funding agency, or closely allying agencies, as trying to satisfy different agendas simultaneously or subsequently is difficult, especially when (re)building CI can be extremely costly. Third, CI design may best be neutral by creating a platform through which competing theories and methodologies can be tested on equal ground.

5. ACKNOWLEDGMENTS

My thanks to Jay Boisseau and Rion Dooley at the Texas Advanced Computing Center for their support of this project.

6. REFERENCES

[1] Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., & Messina, P. 2003. Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. National Science Foundation, Washington, DC. Retrieved December 19, 2006 from http://www.communitytechnology.org/nsf_ci_report/

[2] Bijker, W. E. 1995. Sociohistorical technology studies. In Handbook of Science and Technology Studies, S. Jasanoff, G. E. Markle, J. C. Petersen & T. Pinch, Eds. Sage, London, 229-256.

[3] Deetz, S. 2001. Conceptual foundations. In New Handbook of Organizational Communication: Advances in Theory, Research, and Methods, F. M. Jablin, & L. L. Putnam, Eds. Sage, Thousand Oaks, CA, 3-46.

[4] Hawes, L. 1974. Social collectives as communication: Perspectives on organizational behavior. Quarterly Journal of Speech, 60, 497-502.

[5] Rice, R. E., & Bair, J. H. 1984. New organizational media and productivity. In The New Media: Communication, Research and Technology, R. E. Rice & Associates, Eds. Sage, Beverly Hills, CA, 198-215.

[6] Rogers, E. M. 2003. Diffusion of innovations, 5th ed. Free Press, New York, NY.

[7] Scott, C. R. 1999. Communication technology and group communication. In Handbook of Group Communication Theory and Research, L. R. Frey, D. S. Gouran, & M. S. Poole , Eds. Sage, Thousand Oaks, CA, 432-472.

[8] Scott, C. R. 2003. New communication technologies and teams. In Small Group Communication Theory and Practice: An Anthology, R. Y. Hirokawa, R.S. Cathcart, L. A. Samovar, & L. D. Henman, Eds, 8th ed, Sage, Thousand Oaks, CA, 134-147.

[9] Star, S. L. & Bowker, G. C. 2006: How to infrastructure. In Handbook of New Media: Social Shaping and Social Consequences of ICTs, In L. A. Lievrouw and S. M. Livingstone, Eds. London: Sage, 230-245.

[10] Stewart, C. 2007. Indiana University cyberinfrastructure newsletter. Retrieved November 1, 2008 from http://racinfo.indiana.edu/newsletter/archives/2007-03.shtml.

Information Doesn't Want to Be Free: The Irreducible Costs of Information

Rebecca Crist

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 E. Daniel Street, MC-493
Champaign, IL 61820-6211 USA
+1-217-244-2249
rcrist@illinois.edu

M. Kathleen Kern
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 E. Daniel Street, MC-493
Champaign, IL 61820-6211 USA
+1-217-244-3604
katkern@illinois.edu

ABSTRACT

Since first being pronounced in 1984, the phrase "information wants to be free" has echoed through the corridors of information management as a rallying cry, an aspiration, and a fact. But can it be true?

The champions of Web 2.0 propose that we live in an age when anyone can be a publisher. Wiki media and institutional repositories allow authors to publish freely and allow users free access. No-fee services are widely available for blogs, instantaneous status updates, and widespread dissemination of even the most trivial communications, to even the most micro-scale niche audiences. Have we finally reached an age when information can really be free?

The costs of producing information are undeniable. The funding models for publishing it are clearly shifting, and the business of scholarly communication is being revolutionized—increasingly corporatized on the one hand, and in search of viable open-source options on the other. It is critical that information professionals examine the economic issues of information distribution during this transformation process to ensure greatest success for reliable, stable, accessible, and equitable dissemination of scholarly knowledge.

This poster examines the real financial costs of producing and providing information. Open publishing formats depend on actual people doing actual work—often unseen and unacknowledged. The chain of funding and support for academic research is complicated and hierarchical; researchers providing knowledge "for free" as a service to their field are typically drawing a salary from the university, which is itself funded by taxpayers and donors. Research is funded by corporate grants and academic salaries. Yet would deducting the acknowledged costs of research and publication create a free-information model? We conclude that it would not. By comparing the costs of traditional publishing with their open-access "cost-free" alternatives, we find that we still cannot eliminate the actual expenses of publishing scholarly research. Development costs for open-access architecture remain legitimate and undissolvable costs, regardless of access philosophy. When research is posted, hosted, edited, and peer-reviewed voluntarily by an open scholastic community, we may remove those fiscal expenses from the publication equation. What remains—the invisible work of infrastructure, architecture, testing, implementation, and the like—must still be funded. From software developers to hardware assembly line workers, the work of non-academic participants cannot be deducted from the university funding model.

Additionally, the open-publishing options that require unpaid scholarly services still depend on a variation of the traditional for-compensation economic model: We trade the commodity of time for the currencies of information exchange, academic prestige, and publication credit. We rely increasingly on unpaid academic labor for editing and vetting scholarly output. These "free" services are not without costs. Mindful of the time and skill put into this work, the community must examine the ethics of using "free" labor. What are the guarantors of quality in a no-compensation work-place? How much are faculty willing to contribute beyond being on reviewing boards? Will they, and can they, perform the tasks that hired professional copy editors, indexers, proofreaders, and catalogers once performed? What price do faculty pay for the increased pressure to donate extracurricular labor?

Acknowledging the real costs of open-access scholarly publication will better enable information professionals to seek reasonable long-term solutions to the problems of information distribution. Ignoring these costs leaves information vulnerable to corporate influence and to obsolescence. Moreover, suggesting that information wants to be free disregards the morality of compensating all players—not just the professors and journal editors, but the code scripters and the server masters and the third-shift IT techs—for real work provided. In the end, information may want to be free, but it is unlikely that it ever will be.

Categories and Subject Descriptors

K.6.0 [**Management of Computing and Information Systems, General**]: Economics

General terms

Economics; Human Factors.

Topics

Information economics, Scholarly and scientific communication

Keywords

Publishing models; information economics; open access

Factors Influencing the Adoption of Social Media in the Perspective of Information Needs

Youngseek Kim
Syracuse University
School of Information Studies
221 Hinds Hall
Syracuse, NY 13244
ykim58@syr.edu

Minjae Kim
Syracuse University
School of Information Studies
327 Hinds Hall
Syracuse, NY 13244
mkim46@syr.edu

Kyungseek Kim
Syracuse University
School of Information Studies
327 Hinds Hall
Syracuse, NY 13244
kkim47@syr.edu

ABSTRACT

This study focuses on the factors which affect individual users' adoption of social media in the perspective of their information needs. Social media are the emerging digital communication channels which provide information sharing grounds by helping users distribute and consume information. We introduced both adoption and gratification approaches of IT/IS adoption research by considering individual users and their information needs. For this empirical study, we reviewed previous literatures and their theoretical frameworks in regards to individual users' adoption of IT/IS. Then, we developed the social media adoption model by including significant constructs including perceived usefulness, perceived ease of use, perceived enjoyment, and intention to use from TAM and its extended models. In addition, we also included two more constructs including social influence and personal innovativeness as moderating factors. Finally, we will empirically test our model by using a survey method in order to understand individual users' adoption of social media.

Keywords

Social Media, IT/IS Adoption, Information Needs, Gratification Approach, TAM, Innovation Diffusion Theory

1. INTRODUCTION

Social media are the emerging digital communication channels which create a user-oriented information sharing ground where any people can generate or subscribe information content as both information provider and consumer. Social media have steadily increased among adult American Internet-users according to Pew Internet Research [6]. And they have become an important source of information for online users. Agichtein and his colleagues found that social media provide a rich variety of information sources: in addition to the content itself, there is a wide array of non-content information available [2].

The social media can be considered as an information system which users distribute and consume information. In the Information Systems (IS) field, individual users' IT/IS adoption has been studied in organizational contexts. Also, researchers already have empirically examined the determinants of IT/IS usage [5, 7, 9, 10]. However, there are just few studies why individual users adopt social media for their own usage, especially to satisfy their information needs. This study is interested in individual users' adoption of social media in the perspective of their information needs.

As a starting point of this research, relevant literatures including theoretical background were reviewed. Based on the review of existing literature, we developed a research framework to guide a survey study, which will be used to collect data for the research questions. Following the research framework, the survey questionnaire will be developed and tested with pilot study sample. Then, actual data will be collected through an online and offline survey of the general Internet users. Factor analysis and multiple regression analysis will be conducted based on the collected data. Finally, we will interpret and discuss about the data analysis based on our research framework.

2. THEORETICAL BACKGROUND OF IT/IS ADOPTION

There are various theoretical models which explain individuals' IT/IS adoption. According to Verkasalo, the adoption research can generally be divided into four categories including diffusion research (market focus), adoption approach (individual user focus), gratification research (needs of users focus), and domestication research (consequence of adoption focus) [12]. In this study, we will focus on both gratification research and adoption approach by considering individual users and their information needs.

The Theory of Reasoned Action (TRA) and Theory of Planed Behavior (TPB) provide the basic theoretical framework for understanding users' innovation adoption [3, 4]. The TRA and TPB have influenced the Technology Acceptance Model (TAM) and its extended models. Davis presented the TAM to explain the determinants of user acceptance of a wide range of end-user computing technologies [5]. In the TAM, both perceived usefulness and perceived ease of use affect the intention to use, which eventually influences the real usage behavior. Later, the TAM model has been developed by adding determinants which affect perceived usefulness and perceived ease of use.

Venkatesh and Davis enhanced the TAM to TAM2, which provides a detailed account of the key forces underlying judgments of perceived usefulness [10]. TAM2 was expanded to include the impacts of three interrelated social forces including

subjective norm, voluntariness, and image as determinants of perceived usefulness. Later, the Unified Theory of Acceptance and Use of Technology model (UTAUT) was developed by Venkatesh and colleagues [11]. UTAUT provides a refined view of how the determinants of intention and behavior evolve over time and assumes that there are three direct determinants of intention to use (performance expectancy, effort expectancy, and social influence) and two direct determinants of usage behavior (intention and facilitating conditions) [11].

For this study, we introduce the innovation diffusion theory to explain individual's social media adoption behavior [8]. Drawing upon Rogers' theory of the diffusion of innovations, Agarwal and Prasad described personal innovativeness as the willingness of an individual to try out any new information technology [1]. They added this individual difference variable as a new construct to Davis's original TAM model and hypothesized that individuals with higher levels of personal innovativeness are expected to develop more positive perceptions about the innovation in terms of advantage, ease of use, compatibility, etc. and have more positive intentions toward use of a new IT/IS [1].

3. RESEARCH FRAMEWORK

Based on the theoretical background and literature review, we can identify major constructs which may affect individual users' intention to adopt social media to satisfy their information needs. We developed individuals' social media adoption model by including significant elements from TAM model. If users believe that social media are complicated to use for their information purposes, they may not want to use the social media in order to distribute or acquire information. People may want to use social media if they believe the social media are useful by satisfying users' information needs. Also, since the social media provide various entertainment-related functions through their information ground, the enjoyment of social media would be a good reason why people want to adopt social media for their information related entertainment purposes.

In addition to the direct factors to the intention of individuals' adoption of social media including perceived usefulness, perceived ease of use, and perceived enjoyment, there are moderating factors which affect the individuals' intention to adopt social media. These moderating factors include social influences and personal innovativeness. In this study, social influence refers to perceived pressures from other people and media to make or not to make a certain behavioral decision. Social influence can play a significant role in affecting users' adoption of social media for their information sharing behaviors. Also, Personal innovativeness is included in this research since it has been expected to influence individuals' intention to adopt social media. Personal innovativeness is defined as the willingness of an individual to try out any new information systems [1]. We believe that these moderating factors would directly affect the intention to use social media, and also they would indirectly affect the intention to use social media services by influencing the direct factors including perceived usefulness, perceived ease of use, and perceived enjoyment.

4. RESEARCH DESIGN & METHOD

We will use a survey method to conduct this research. To examine the conceptual constructs and hypothesized relationships, the survey questionnaire will be developed for the adoption of social media in the perspective of information needs. The survey items will be brought from previous studies and modified for this research. Then, we will create an actual survey questionnaire by adjusting the measures of the constructs according to the interviews with actual social media users. For the questionnaire, all the items from previous studies will be modified to make them relevant to the social media usage under the information behavior context.

The survey data will be collected through online survey. The population in this study focuses on the Internet users in the U.S., and the sample population which we will collect is targeted 250 participants. The sample would be collected by using a commercial survey response website. Since college students are more the early adopters in IT/IS adoption, it would be appropriate not to use the college students in this survey. Later, we can generalize the results of this research to the population of the general Internet users.

The empirical data will be gathered to test our research model. The questionnaire will consist of research introduction and purpose, specific questions to measure the constructs, and respondents' demographic information. Each measurement for the constructs including perceived usefulness, perceived ease of use, perceived enjoyment, social influence, personal innovativeness, and intention to use the social media will be collected.

After we get the final data set, we will clean the survey data set by removing invalid data. The descriptive data analysis will show some demographic information about the respondents. In regard to statistical analyses, factor analysis and multiple regression analysis will be conducted based on the collected data. For the measurements of six different variables, Principal Factor Analysis can be used with VARIMAX rotation (if needed). Correlation analysis will be used to see the relationships among all the research variables.

5. CONCLUSION

The purpose of this study is to understand the factors which influence individuals' adoption of social media in the perspective of their information needs. We focus on both gratification research and adoption approach to answer our research questions. In regards to the possible results of this research, first, we believe that perceived usefulness, perceived enjoyment, and social influence are the important determinants of the social media adoption. Second, we also believe that perceived enjoyment rather than perceived usefulness and perceived ease of use may have a greater impact on individual users' intention to adopt social media. Third, we think that individual users' perception of usefulness, ease of use, and enjoyment may be significantly attributed to social influence from users' social network. Lastly, individual users' perception of usefulness, ease of use, and enjoyment may be significantly attributed to personal innovativeness. The analysis of the survey results will help us validate the new model and understand individual users' adoption of social media.

6. REFERENCES

[1] Agarwal, R., and Prasad, J. "A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology," Information Systems Research (9:2) 1998, pp 204-215.

[2] Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008, February 11-12). Finding High-Quality Content in

Social Media. Paper presented at the ACM International Conference on Web Search and Data Mining, Palo Alto, California, USA.

[3] Ajzen, I. "The Theory of Planned Behavior," Organizational Behavior and Human Decision Process (52:2) 1991, pp 179-211.

[4] Ajzen, I., and Fishbein, M. Uncerstanding Attitudes and Predicting Social Behavior Prentice-Hall, Englewood Cliffs, NJ, 1980.

[5] Davis, F.D. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance in Information Technology," MIS Quarterly (13:3) 1989, pp 319-340.

[6] Lenhart, A., & Fox, S. (2009, February 12). Twitter and status updating. Pew Internet & American Life Project Retrieved October 2, 2009, from http://www.pewinternet.org/Reports/2009/Twitter-and-status-updating.aspx

[7] Moore, G.C., and Benbasat, I. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," Information Systems Research (2:3) 1991, pp 192-222.

[8] Rogers, E.M. Diffusion of Innovations (5th Edition) The Free Press, New York, NY, 2003.

[9] Taylor, S., and Todd, P.A. "Understanding Information Technology Usage: A Test of Competing Models," Information Systems Research (6:2) 1995, pp 144-176.

[10] Venkatesh, V., and Davis, F.D. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," Management Science (46:2) 2000, pp 186-204.

[11] Venkatesh, V., Morris, M.G., Davis, G.B., and Davis, F.D. "User Acceptance of Information Technology: Toward a Unified View," MIS Quarterly (27:3) 2003, pp 425-478.

[12] Verkasalo, H. "Dynamics of Mobile Service Adoption," International Journal of E-Business Research (4:3) 2008, pp 40-63.

Toward a Theoretical Framework for Digital Age Information Behavior of Youth

Kyungwon Koh
School of Library and Information Studies, Florida State University
Shores Building, 142 Collegiate Loop
Tallahassee, FL 32306-21001
1-850-766-6452
kwk05@fsu.edu

Keywords
Information behavior, youth, digital age, Radical Change theory

1. BACKGROUND
Recent studies in various disciplines—such as Education, Communication, Media Studies, Psychology, Law, Business, Sociology, and Library and Information Studies [LIS]—suggest that today's young people think, learn, socialize, shape identity, and seek information differently in this digital information age, the era of Web 2.0 and of participatory culture [1]. Several terms are applied to describe members of this unique generation who are growing up immersed in digital technologies from the start of their lives, including the Net Generation, Generation M (M for Media), 21st century learners/ students, Digital Natives, and digital age youth. In general, researchers define these groups as including individuals who were born after a certain year, ranging from 1978 -1989.

2. PROBLEM STATEMENT
The study assumes that contemporary youth's predominant engagement in digital media culture influences many aspects of their lives and results in some fundamental changes in their information behavior from the social constructivist point of view. For example, they use multiple media sources to seek information, are exposed to an increased array of information with diverse perspectives, actively create information, and exhibit interactive, nonlinear, and collaborative information behaviors. Since the attributes are quite different from those of traditional information behaviors, it is important to understand digital age youth's approaches to information seeking and provide information services that match their new characteristics and patterns.

Yet, few theoretical frameworks and empirical studies exist to identify and explain changes in digital age youth information behavior in iField. The most recent national guidelines and standards for school libraries reflect the momentous changes for learners in the digital age, addressing multiple literacies, a continuing expansion of information and the social nature of learning facilitated by digital technology. However, youth information behavior models and frameworks have not been updated to manifest the changing notion of information literacy, which has become more complex, as well as the variety of information behaviors in youth everyday life.

3. RADICAL CHANGE THEORY
The theory of Radical Change proposes that three digital age principles—Interactivity, Connectivity, and Access—explain changes in youth information resources and behaviors in the digital age [2]. A typology with three types of changes, each with indicators, operationalizes the theory for identification/ explanation of changes in information resources. The theory, however, has been applied to information behavior of youth without such a typology. Therefore, the proposed study seeks to further develop the theory by establishing a typology (or model) and accompanying variables that address young people's (a) cognitive status, (b) identify/value negotiation and information creation, and (c) social interaction during their interaction with information. The Radical Change theory with a new model resulted from the proposed study will help understand youth information-related activities as a whole and their interrelationships, not just studies of individual tasks or search sessions isolated from the context.

4. RESEARCH PURPOSE
This proposed research aims to understand contemporary young people's information behavior in the digital age based on a solid theoretical and empirical ground. In particular, the exploratory study focuses on new and distinctive behaviors of youth who are engaged with digital media culture. The theoretical and empirical processes of the study result in adding to the original Radical Change theory a model that identifies the key characteristics of youth information behaviors in the digital age.

5. RESEARCH QUESTIONS
The study's research questions are:

1. What are the key characteristics of youth information behavior in the digital age?

1) How do digital age youth seek information and learn? (Cognitive aspects of information seeking)
2) How do digital age youth perceive themselves and others? What part does the creation of information play in negotiating values and forming identities? (Identity, value negotiation; information creation)
3) How do digital age youth access information and seek community? (Information access; collaborative, participatory community seeking)

2. How can Radical Change theory, including a newly added model through the current study, be used to describe, explain, or predict youth information behavior?

6. METHODOLOGY

The study employs a qualitative research design due to its exploratory and holistic nature. A three-phased qualitative methodology design is developed, where each phase must be conducted sequentially because the results of the previous phase will inform and lead to the specific design of the subsequent data collection and analysis. Phases One and Two are intended to answer Research Question 1 (i.e., what are the key characteristics of youth information behavior in the digital age?) and aim at creation and validation of the model of digital age youth information behavior, respectively. Phase Three applies and tests the model in a specific phenomenon of youth formation behavior in this age, focusing on Research Question 2 (i.e., how can Radical Change theory, including the new model of key characteristics, be used to describe, explain, or predict youth information behavior?).

Phase One is a content analysis of existing research studies (creating a model of digital age youth information behavior). Phase Two consists of focus group interviews with a public library Teen Advisory Board, a group of approximately 8 – 10 youth ages 13 – 18, (enhancing credibility of the model from the perspective of youth). Phrase Three will be a document analysis of the ThinkQuest Web database, a collaborative online learning platform in which students work across city, state, and country borders to create learning projects. ThinkQuest students are chosen because they are actively engaged in activities using digital media; this population will provide a valid answer to the question of whether Radical Change Theory can describe, explain, or predict information behavior of digital age youth. Data to be analyzed include each element of the ThinkQuest Project, Competition, and Library, followed by online chat interviews with 3- 6 student participants who collaborate to create a ThinkQuest project. It is expected that Phase Three will demonstrate the applicability of the model to explain a specific phenomena of youth information behavior.

The target population for the entire study is digital age youth in the U.S., 5 – 18 year olds. Students of all ages will be covered in the content analysis of research in Phase One. Phases Two and Three will focus on a sample of older youth, ages 13 – 18, in order to examine more active participation and interaction with digital media with relatively greater autonomy. Collected data will be analyzed using Atlas.ti software for qualitative data analysis.

7. LIMITATION/SCOPE

The study does not conduct a comparative study of information behaviors between older generations and digital age youth. In order to see if some of the noticeable characteristics in today's youth information behaviors are really new, it might be ideal to compare current young people's information behavior and the information behavior of older generations in their childhood, to the extent that such data is available. However, the research does not study youth information behavior historically for comparison purposes, because the goal of the study is to enhance understanding of today's young people and serve them better. Some of the radical change characteristics identified by the study may have also existed to some extent in the past (though they are much more prevalent nowadays), but this fact does not mitigate the importance of understanding such characteristics to provide relevant library and information services for youth in the digital age.

Also, the study focuses on the processes (including cognitive process) or actions while youth engage in information-related activities. Therefore, it is not intended to assess if and how information needs of today's young people have been changing. Assessing information needs of youth in the digital age is beyond the scope of the study.

8. CONCLUSION

Today's young people are engaged in a variety of information activities, and the ways that they interact with information have changed significantly within the past two decades. It is important to understand the changing nature of youth information behavior in order to provide relevant and updated information services for youth that match their unique patterns and approaches to information. Applying the theory of Radical Change, the study suggests that contemporary youth information behavior shows distinct features due to the characteristics of the digital society, which include the digital principles of Interactivity, Connectivity, and Access proposed by the theory. Multiple phases of qualitative research develop and add a new model, which identifies key types and characteristics of digital age youth information behavior, to the Radical Change theory. The most significant scholarly contributions of the proposed research include a theoretical contribution to iField, which provides a new perspective and potential for encouraging research on youth information behavior in the digital age.

9. REFERENCES

[1] Dresang, E. T. and Koh, K. 2009. Radical Change Theory, Youth Information Behavior, and School Libraries. Library Trends. 58, 1, 26-50.

[2] Dresang, E. T. 2005. Radical Change. In Theories of Information Behavior, K. E. Fisher, S. Erdelez, and L. McKechnie, Eds. Information Today, Medford, New Jersey, 398-302.

"Hi! I'm Harvey, A Consent Bot": How Automating The Consent Process In SL Addresses Challenges Of Research Online

Peyina Lin
Information School
University of Washington
Seattle, WA
pl3@uw.edu

Michael B. Eisenberg
Information School
University of Washington
Seattle, WA
mbe@uw.edu

John Marino
Information School
University of Washington
Seattle, WA
marinoj@uw.edu

ABSTRACT
In this paper, we describe the challenges of acquiring informed consent in a virtual environment, Second Life, and describe our employment of an automated consent bot.

Categories and Subject Descriptors
K.4.1 Public Policy Issues – *ethics, human subjects*.

General Terms
Design, Human Factors, Legal Aspects.

Keywords
Informed consent, virtual environments, Second Life, online research

1. INTRODUCTION

Ethical aspects of conducting *research in online environments* (hereafter *research online*) are often framed in terms of challenges associated with informed consent, confidentiality and anonymity [2], [6]. However, in the research we do at the VIBE (Virtual Information Behavior Environments) project of the Information School, University of Washington, we transform these challenges into opportunities to make observation research online more transparent. Specifically, in collaboration with 2b3d.net, our research team automates the informed consent procedure for observations in openly accessible settings in Second Life (SL) through Harvey, a "consent bot." In this article we describe why online research faces various ethical challenges, and how we streamlined the informed consent process and removed potential conflicts that might arise over the debatable "public" quality of behavior in openly accessible spaces in SL.

Second Life (SL) is a 3D virtual environment in which users are represented by a 3D virtual avatar. In SL, users "live" in a 3D virtual space where they shop for clothes, engage in monetary transactions, build friendships, and virtually do almost anything that can be done in our physical lives (via avatars). Spatial ergonomics are built into SL. Volume decreases if your avatar moves away from a sound source and your avatar's perspective of the space changes with its movements. However, due to the extended capabilities users have in SL such as flying, changing appearance, and creating multiple avatars, awareness of the social presence of others can be challenging. Users might not see the avatar standing a few feet behind them or the one flying over their heads, and not be aware of who is tuning into their conversations. In addition, while users may know through SL's Terms of Service, that Linden Lab retains ownership of SL accounts and any related data [4], [5], it is unclear whether users would be notified if Linden Lab were to use their conversations in SL for other purposes. Conscious of these added layers of ambiguity on the privacy and confidentiality of users' conversations, we sought to increase the transparency of our research by ensuring that the informed consent process be systematically approached with every potential participant through the use of a "consent bot"—which automatically detects the presence of an avatar within 20 meters from the bot's position, and informs these avatars of the researcher's presence, research objectives, activities, and enables the users to accept or decline participation on the spot.

iConference'10, February 3–6, 2010, Champaign, Illinois, USA.

2. ETHICAL CHALLENGES OF RESEARCH ONLINE

To better understand how our automated consent bot makes observation research in SL systematic and transparent, it is necessary to elaborate on the challenges of research online. Ethical challenges of research online related to informed consent, confidentiality, and anonymity exist for the following reasons:

1. **Blurred boundaries between the public and private.**

SL is open to all, but is owned by Linden Lab. Its terms of service and community standards prohibit remotely *recording* conversations without permission [1], [3].

2. **The debate over whether online "identities," which are distinct from offline identities, are personal information.**

Observation of "public behavior" is exempt of Federal regulations for the protection of Human Subjects unless the data can be used to identify subjects in a personal way [7]. "Public behavior" refers to "behavior that is apparent to an unconcealed observer, without the use of any special or surreptitious equipment, such as binoculars, special microphones, or recording devices" [7]. Based on this definition, note-taking without identifying avatars in a personal way would be public behavior. However, are SL avatars' names personally identifiable information?

3. **Note-taking while being virtually immersed is challenging.**

Note-taking (not copying and pasting conversations) during observations in a virtual environment can be more challenging than in the physical space: the screen estate may not afford being virtually immersed and note-taking simultaneously.

4. **Continuous informed consent disrupts the flow of naturalistic observations.**

Copying and pasting the local chat conversations is a form of recording, and participants' consents need to be obtained to meet Linden Lab and IRB requirements. However, in open spaces where avatars come and go, how can we obtain informed consent continuously without disrupting the flow of naturalistic observations?

3. HARVEY THE "CONSENT BOT": HOW WE STREAMLINED THE INFORMED CONSENT PROCESS

Figure 1 illustrates the logical path that our consent bot, Harvey, takes to inform potential observation participants that a researcher is conducting observations in the SL space in which they find themselves, and subsequently obtains their consent to textually record their local chat conversations.

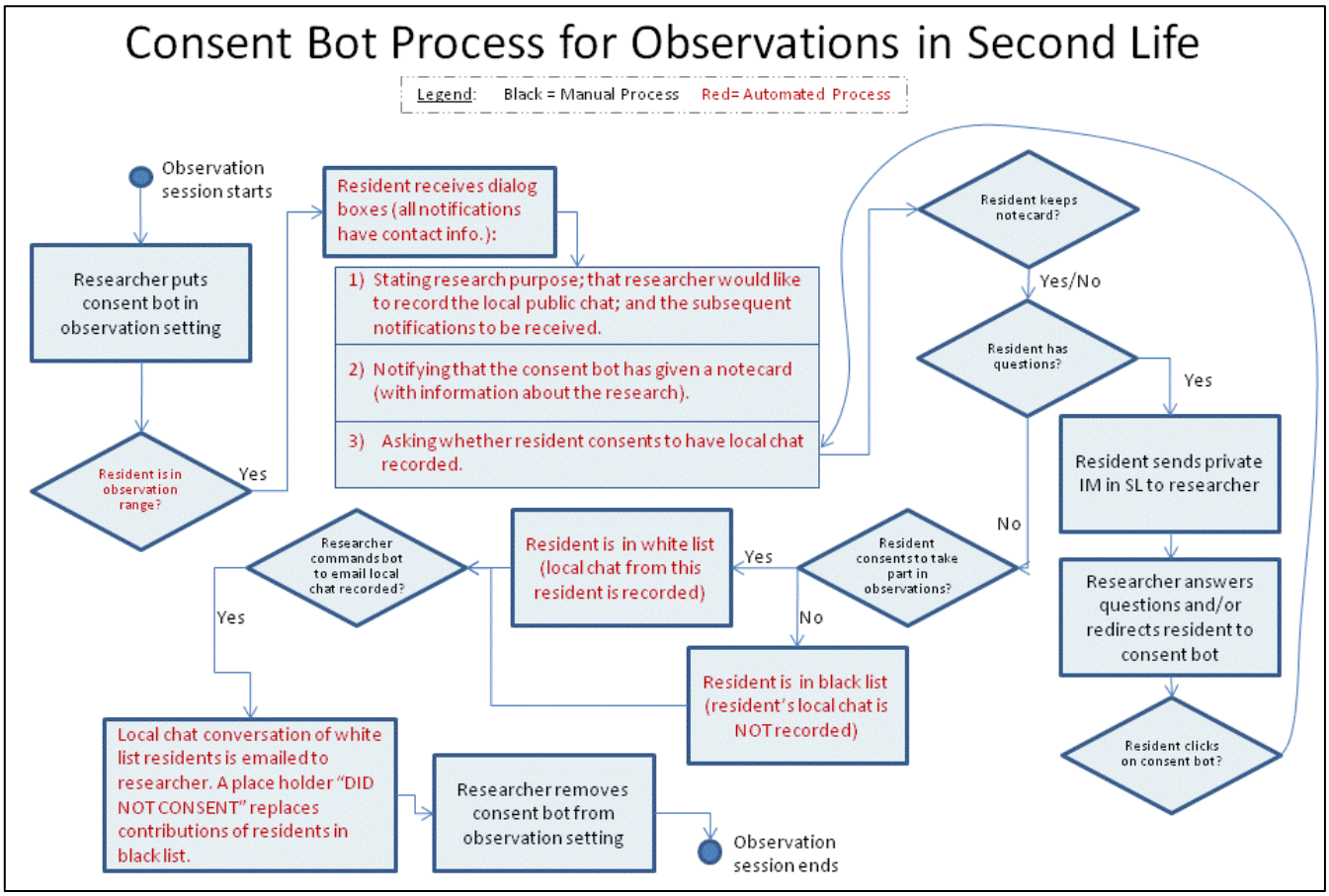


Figure 1. Consent Bot Process for Observations in Second Life

As discussed, our research team was aware that note-taking while simultaneously being immersed in virtual space would be difficult, and the temptation to copy and paste the conversations would always be there. Therefore, by choosing to textually record SL's local chat conversations during times when a researcher was virtually present in an openly accessible SL space, we were clearly positioning ourselves. We could not treat the behavior in SL's openly accessible spaces as "public behavior", in terms of Human Subject's definition. We were also required to obtain explicit consent under Linden Lab's Terms of Service. However, we were faced with the challenge of how to obtain continuous informed consent without disrupting the quality of naturalistic observations. We saw this challenge as an opportunity. We took advantage of the virtual extended capabilities that can be programmed into objects in SL, and in collaboration with 2b3d.net, developed a consent bot named Harvey.

The informed consent process starts when the researcher enters a setting selected for observation and places Harvey, the consent bot object, in the setting. Once out in the field, Harvey sends out a notification in the form of a SL dialog box to any avatar that is within 20 meters of distance (dialog boxes are SL's standard notification system). The dialog box summarizes in one sentence why the avatar is receiving the notification—because "researchers from the VIBE team would like to observe you and record your local chat to help understand patterns of information behavior and improve user experiences in virtual worlds." It is important to note that people tend to not read through long pieces of text. Thus, it was crucial to include the gist of what the potential participant would be consenting to in one short sentence at the beginning of the dialog box. The dialog box also provides the researcher's SL contact information, and summarizes in bullets the steps that the potential participant would go through to complete the informed consent process.

Potential participants would click OK to proceed with the process, or click on "ignore" to ignore the dialog box. If they ignored the

first informational dialog box, their conversations would not be recorded.

For those who clicked OK to the first informational dialog box, they would receive a second dialog box notifying them that the researcher would like to give them a notecard. This notecard contains more information about the research observations. Potential participants would "keep" or "discard" the notecard.

If potential participants kept the notecard, this notecard would pop up, and participants could read more about the research study, including the standard information that is included in a Human Subjects consent form such as purpose of study, confidentiality, and risks. Regardless of whether participants keep or discard the notecard, potential participants receive a third dialog box which asks whether they consent to having their local chat recorded while a researcher conducts observations. This dialog box summarizes again the gist of what the potential participants would be consenting to, in case they didn't read the notecard. The dialog box also provides a URL where participants can read more information about the research activities, and the researchers contact information in SL.

Those who consent (or click yes) go to a white list. Only the local chat conversations of those in the white list are recorded. A "did not consent" text appears in lieu of the textual contribution of those who did not consent or who ignored the consent bot.

Harvey is also designed to take care of those who change their minds or came after the researcher put Harvey in the observation location. In parallel to using Harvey, the researchers also use other non-automated techniques to make potential participants aware of the research activities including "wearing" a researcher sign above their heads and displaying a clearly visible sign that notifies passers-by of the observation activities.

4. CONCLUSION

Conducting research in virtual environments can be challenging when it comes to addressing the requirements of IRB review boards, corporate owners of the virtual environment, and users' expectations. However, the programmable nature of virtual worlds enables the creation of processes that can relieve these concerns. From a user's standpoint, receiving a few informational dialog boxes pushed by the researcher is a small price to pay for being transparently informed about how the user's interactions in that virtual space may be used. We hope that the design of Harvey the consent bot acts as a model of informed consent process in SL and other virtual worlds.

5. ACKNOWLEDGMENTS

Thanks to the John D. and Catherine T. MacArthur Foundation for funding this research, and especially Kathy Im for enabling this work. Thanks to current and former members of the VIBE research team, Malik Hussain and Eric Meyers for their contributions to the research project. Special thanks to Randy Hinrichs, Janice Cowsert Hinrichs, and their 2b3d.net for the development of Harvey and VIBE's virtual space in SL and to the tech-support team of the University of Washington's iSchool.

6. REFERENCES

[1] Conduct by Users of Second Life, in Second Life's Term of Service. Online: http://secondlife.com/corporate/tos.php

[2] Esenbach, Gunther and Till, James E. Ethical issues in qualitative research on internet communities. *BMJ*. Accessed October 11, 2009 online: http://www.bmj.com/cgi/content/short/323/7321/1103

[3] Disclosure, point 4, in Second Life's Community Standards. Online: http://secondlife.com/corporate/cs.php

[4] License Terms and Other Intellectual Property Terms, point 3.2, in Second Life's Terms of Service. Online: http://secondlife.com/corporate/tos.php

[5] License Terms and Other Intellectual Property Terms, point 3.3, in Second Life's Terms of Service, online: http://secondlife.com/corporate/tos.php

[6] TAYLOR, T. 1999. Life in virtual worlds: Plural existence, multimodalities, and other online research challenges. *Am Behav Sci* 43, 436-449.

[7] University of Washington Human Subjects Manual, section II-C. Categories of Research—No Risk Research (Exemptions). Online: http://www.washington.edu/research/hsd/hsdman2.html

Social Inclusion in High School: A baseline behavioral study to inform the design of pro-diversity technology

Peyina Lin
The Information School
University of Washington
pl3@uw.edu

ABSTRACT
While social groups and social technologies play important roles in the lives of American high school youth, it is not well understood how these two aspects structure opportunities for civic participation and shape social boundaries. This study reports work-to-date on how group and friendship structures affect high school students' extracurricular and civic participation. Future study phases aim at describing how social technologies, like Facebook, MySpace, instant messaging, and mobile technologies are used in friendship and extracurricular contexts, and whether social technologies affect the interrelationships between friendship structures and extracurricular participation. Expected contributions include revealing the potential of social technologies to restructure friendships and opportunities for participation from diverse youth.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human Factors– *effects of social structure on interaction.* H.5.3 [**Group and Organization Interfaces**]: Collaborative computing—technology for social inclusion.

General Terms
Human Factors.

Keywords
Social technologies, social media, high school social groups, friendship structures, extracurricular participation, civic participation, social inclusion, diversity.

1. INTRODUCTION
Extracurricular activities in high schools often aim to give students the opportunity to take part in activities of interest, develop sense of community and expanded sense of self. However, high school students do not always have a choice. Social structures such as previous experience, performance, skills, popularity, ethnicity, teacher preference, socio-economic status, and existing friendships can systematically predispose students to opportunities for participation [13], [10]. This can have social implications, particularly in an area like civic participation. Civic participation in high school means doing things such as volunteering for a charitable cause, speaking out on public and community issues, and representing the student body [1]. Researchers suggest that civically active high school youth are more likely to be civically active adults [8]. Therefore, ensuring that the next generation participates in our society with efficacy is important, and requires our society to move beyond systematic structural predispositions to enable opportunities for diverse youth.

Previous research has found a weak association between high school students' friends and their civic attitudes [9], [14] yet this work has ignored the socio-technical context in which adolescents make friends. Within schools, social groups (recognizable via labels such as "jocks" or "populars") embed students in a social status hierarchy ([2],[3],[5],[6],[13]) that may structure extracurricular opportunities. Technologically (and socially), adolescents use social technologies (such as social network sites, instant messaging, and the mobile phone) for inclusion and acceptance [4], and group definition and boundary setting [11]; behaviors which are relevant to the understanding of the opportunities afforded by social groups. However, little is known about how the structure of friendships (e.g., the status of social groups, common activities, and technologically mediated interactions) predispose high school students to participation in the student government and other service activities. For example, are those more central in high school social life more likely to use technologies in certain ways? Do students more connected to others have more student government leadership opportunities?

Drawing from social structural theory, as well as research on high school culture, adolescent reference groups, and adolescents' use of social technologies, this mixed methods study examines: a) whether friendship structures affect (or are associated with) high school students' extracurricular and civic participation; b) how social technologies are used in friendship and extracurricular contexts; and c) whether social technologies affect the interrelationship between friendship structures and extracurricular participation. The overarching research question is: **To what extent do social technologies reproduce the interrelationships between friendship structures and extracurricular participation?**

This article describes the work-to-date, impressions from the field, expected contributions, and future phases of this study.

2. WORK TO DATE
The study is set in two high schools in the suburbs nearby Seattle, WA: The high schools have comparable demographics and the same technology use policies, but one has a higher SES.

Instrument feasibility--pilot study one (08/2008): A pilot study was conducted with 4 adolescents (14 to 15 years old) to test the viability of the measurements for social groups, theoretical constructs, and friendship structure. The pilot study showed that the measurements were viable. While no pattern of technology use was established with only 4 respondents, *homophily*, or the

likelihood to bond or associate with others like oneself [9], was evident. ¾ of participants' top 3 friends were of the same gender; and ¾ of participants' friends were in the same social group.

Teacher interviews (08/27 – 09/18, 2009)**:** Interviews were conducted with 10 teachers from both schools to learn about the school's course structure and develop a clustering sampling plan. Teachers also helped to localize the list of social groups [2] to be used in the student questionnaire.

Pilot study two (09/17 – 23, 2009): The revised questionnaire from pilot study one was refined with data from 20 students from both schools.

Student questionnaire (10/01 – present): Cluster sampling was used in both high schools (grades 10 through 12). 7 to 12 English classes (approximately 400 to 500 students) per school were sampled from grades 11 and 12. Pilot study two showed that some 10^{th} graders could not answer questions about social groups and friendships in school at the start of the school year; thus 10^{th} graders were not included in the initial sampling. To date, an average of 130 students per school have taken the online questionnaire. A total of 300 students are expected to take the questionnaire. The questionnaire asks about family background, technology ownership and use, social group and activity participation, and use of social technologies with different people (close friends, students from the same social group, and people in their civic group). The questionnaire provides the basis for analyzing structural characteristics of students' social relationships and use of social technologies. In order to answer the research question meaningfully, reporting of the questionnaire data is withheld until its collection is completed in January of 2010.

Informal observations (08/27 – present): 37 trips to both schools, each lasting an average of 2 hours have taken place. During the time I spent in the schools, unintended interactions were initiated by teachers and students. As described below, these interactions generated naturalistic observations which appear to validate the study's expected contributions.

3. IMPRESSIONS FROM THE FIELD

Here I describe impressions and preliminary analysis from one high school. Teachers and students here claim that the school does not have clicks or is much more accepting of difference. While this appears to be the case when compared to the other high school, the status of social groups is still evident. For example, during their homecoming assembly, the homecoming queen and king were a tall cheerleader and handsome football player. They were glorified in the homecoming assembly and at the homecoming football game. Coincidentally, cheerleading and football were the two most mentioned extracurricular activities by questionnaire participants (n=147) to describe the "popular" and "jocks" social group. These two social groups were nominated the most frequently (53 and 17) as the social group to be at the center of school social life. Teachers also believe that officers of the Associated Student Body (i.e., the student government) are at the center of school events. This also coincides with the next most representative activities for the "popular" and "jock" groups mentioned in the questionnaire: ASB and spirit day. This suggests that popularity, an indicator of social status, is related to being a student representative in the ASB. If being a student representative is the quintessential form of high school civic/political participation, and civic participation in high school determines participation in adulthood, are the non-popular kids doomed to not be civically/politically involved adults? While not suggesting a positive answer, the data raises the need to devise civic opportunities for diverse youth.

4. WORK IN PROGRESS/FUTURE PHASES

Data cleaning is taking place in order to analyze the patterns of personal and technologically mediated interactions with best friends, and with people in one's social and extracurricular groups. Once the collection of student questionnaire data is complete, questionnaire data will be analyzed to address situational hypotheses pertaining the a) the interrelationships between the status of school-based social groups and students' choice of extracurricular activities; (b) the overlap between friendships and extracurricular participation; and c) how social technologies affect the patterns observed in (a) and (b).

Future study phases (01 – 05, 2010): Interviews and observations will help to interpret findings from the questionnaire. 12-20 students from 4-5 activity groups, and some not active in any groups, will participate in interviews. Critical sampling will be used to select students who most meaningfully illustrate theoretical concepts. Students will demonstrate how they use different technologies to interact with different groups of people. Students' behavior inside and outside their social groups, during free time, activities time, and school events will be observed.

5. EXPECTED CONTRIBUTIONS

This study will report the theoretically based situations in which social technologies are more likely to make a difference on students' opportunities for extracurricular and civic participation. For example, what social technologies are more likely to make a difference on the participation of students of medium to low social status groups and with high interaction with school groups? What are the theoretically based conditions in which heavy use of social technologies is more likely to have positive effects on exposure to social others like or unlike oneself? The study intends to take these behavioral findings one step further, and generate design implications for pro-diversity and pro-inclusion technology that motivates students to interact across group boundaries. This is of particular interest in high school settings because general purpose email and social network utilities are blocked. If, however, social technologies have a pro-inclusion and pro-diversity potential, schools may have to reconsider their technology policies.

6. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. SES-0927291. Thanks to the school district, schools, teachers, and students who participated in the study. Thanks to my advisor Michael B. Eisenberg for his continued advice support; and to committee members Kate Stovel and David McDonald for their constructive criticism.

7. REFERENCES

[1] Andolina, M. W., Keeter, S., Zukin, C., & Jenkins, K. 2003. *A guide to the index of civic and political engagement.* The Center for Information & Research on Civic Learning &

Engagement. Retrieved December 10, 2006, from http://www.civicyouth.org/PopUps/IndexGuide.pdf

[2] Brown, B. B., & Lohr, M. J. 1987. Peer-group affiliation and adolescence self-esteem: An integration of ego-identity and symbolic-interaction theories [Electronic Version]. *J Pers Soc Psychol, 52*, 47-55. Retrieved September 20, 2007.

[3] Brown, B. B., & Steinberg, L. 1990. Skirting the "brain-nerd" connection. Academic achievement and social acceptance [Electronic Version]. *The Education Digest, 55*, 57-60. Retrieved September 20, 2007, from http://proquest.umi.com/

[4] Boyd, D. 2008. Why Youth (heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life." *Youth, Identity, and Digital Media.* Edited by David Buckingham. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. (pp. 119–142). Cambridge, MA: The MIT Press. doi: 10.1162/dmal.9780262524834.119

[5] Coleman, J. S. 1961/1981. *The Adolescent Society*. Westport, Connecticut: Greenwood Press, Publishers.

[6] Garner, R., Bootcheck, J., Lorr, M., & Rauch, K. 2006. The adolescent society revisited: Cultures, crowds, climates, and status structures in seven secondary schools [Electronic Version]. *J Youth Adolescence*, 35, 1023-1035. Retrieved May 14, 2006 from http://www.springerlink.com/

[7] Haythornthwaite, C., & Wellman, B. 1998. Work, Friendship, and Media Use for Information Exchange in a Networked Organization [Electronic Version]. *J Am Soc Inf Sci 49*, 1101-1114. Retrieved January 2, 2007, from DOI: 10.1002/(SICI)1097-4571(1998)49:12<1101::AID-ASI6>3.0.CO;2-Z

[8] Jennings, M. K., & Niemi, R. B. 1974. *The Political Character of Adolescence*. Princeton: Princeton University Press

[9] Kandel, D. B. 1978. Homophily, selection, and socialization in adolescent friendships [Electronic Version]. *Am J Sociology*, 84, 427-436. Retrieved Feb. 5, 2007 from http://www.jstor.org.

[10] Loder, T. L., & Hirsch, B. J. 2003. Inner-city youth development organizations: The salience of peer ties among early adolescent girls. *Appl Dev Sci, 7*(1), 2-12.

[11] Ling, R., & Yttri, B. 2002. Hyper-coordination via mobile phones in Norway. In J. E. Katz & M. Aakhus (Eds.), *Perpetual Contact* (pp. 139-169). Cambridge, UK: Cambridge University Press.

[12] London School of Economics, Centre for Civil Society (2004). *What is a civil society?* Retrieved December 15, 2006, from http://www.lse.ac.uk/collections/CCS/what_is_civil_society.htm

[13] McNeal, R. 1998. High school extracurricular activities: Closed structures and stratifying patterns of participation. *J Educ Res, 91*(3), 183-191.

[14] Sebert, S. K., Jennings, M. K., & Niemi, R. G. (1974). The political texture of peer groups. In M. K. Jennings & R. G. Niemi (Eds.), *The Political Character of Adolescence* (pp. 229-248). Princeton, New Jersey: Princeton University Press.

Analysis of Query Reformulation Types on Different Search Tasks

Chang Liu, Jacek Gwizdka, Nicholas J. Belkin
School of Communication and Information, Rutgers University,
4 Huntington Street, New Brunswick, NJ 08901, USA
changl@eden.rutgers.edu, iConf2010@gwizdka.com, belkin@rutgers.edu

Abstract

This study examined the frequency of query reformulation types in task-driven experiment, and how it is related to task type and individual differences among users. Our results indicated that Specialization and Word Substitution were the two most frequent query reformulation types. In addition, the type of search task had a significant effect on the type of query reformulation.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval–query formulation

General Terms

Measurement, Design, Experimentation, Human Factors

Keywords

Query reformulation, search task

1. INTRODUCTION AND RESEARCH QUESTIONS

When people are searching information online, they often have difficulties in choosing correct words to represent their information problems [2] and, typically, they have to modify or reformulate their previous queries to find useful information and accomplish their information goals. Previous studies on query logs in some major search engines have proposed classifications of reformulation types ([8]; [10]) and how users' click-through behaviors varied by different types of reformulation ([7]). But the method of using query logs cannot provide the information goals behind the queries; without such context, the reason why people are using specific types of query reformulation remains unknown.

In prior work, researchers have examined some contextual factors on query strategies. For example, Hsieh-Yee [6] found the influence of both search experience and subject knowledge on the search tactics; Aula [1] suggested that the levels of experience in using Web search engines and computers all made a difference in query formulation behaviors. Wildemuth [12] showed the effects of domain knowledge on search tactic formulation.

Ford, Miller & Moss [3] found the relationships between search strategies and some individual differences (e.g. cognitive/demographic individual differences, study approaches, and perceptions/approaches). In addition, their results also indicated that the changes of search strategies were seemingly in response to the increasing task difficulty.

In the current study, we will examine the usage of reformulation types when users are searching for some pre-designed search tasks, which can provide us their context information. The research questions for this study are:

1) What query reformulation types are frequently used by searchers?

2) Whether and how does the frequency of each query reformulation type vary among different search tasks?

3) Whether and how does the frequency of each query reformulation type vary between people characterized by different cognitive abilities?

In order to examine these research questions, we first constructed a taxonomy of query reformulation strategies adopted by users; and then identified the types of query reformulation during the process of their searching for search tasks. In this study, the frequency of each reformulation type is the dependent variable, and the task features and individual cognitive ability levels are independent variables.

2. METHOD

We conducted a task-driven, web-based information search experiment. Participants were asked to conduct 6 search tasks of different type and structure on the English Wikipedia. Each study session was one and a half to two hours long, and user interaction with computer was logged. 48 participants (17 females and 31 males) participated in this study, and their average age was 27 years.

Tasks

We used 12 search tasks in this study. Eight tasks out of the 12 were created by Toms et al. [10] for INEX 2007 [9] and 4 simple search tasks were created by us. During each study session, participant performed 6 tasks of differing type and structure (Table 1). For each task, participant was able to choose between two tasks of the same type and structure but on different topics.

Among all 12 search tasks, two types of search tasks were identified: Fact Finding (FF) and Information Gathering (IG). The goal of a fact finding task is to find one or more specific pieces of information (e.g., name of a person or an organization, product information, a numerical value; a date). The goal of an information gathering task is to collect several pieces of information about a given topic.

The tasks were also divided into three categories that depended on the structure of the underlying information need: Simple (S), where the information need is satisfied by a single piece of information; Hierarchical (H), where the information need is satisfied by finding multiple characteristics of a single concept; Parallel (P), where the information need is satisfied by finding multiple concepts that exist at the same level in a conceptual

hierarchy ([10]). By definition, there were five possible combinations of task types and structures: FF-S, FF-H, FF-P, IG-H, and IG-P. In addition, three levels of objective difficulty were assigned to these tasks, as shown in Table 1. We also asked participants to evaluate the difficulty of each task after they finished searching (level: easy, med, difficulty), which was subjective difficulty.

Table 1. Variable facet values for the search tasks

Task id	Task Type	Task Structure	Combined Task Type	Objective Difficulty	Subjective Difficulty
1 2	FF (Fact Finding)	Simple (S)	FF-S	Easy	Easy Med Difficult **(self-rating by users)**
3 4					
5 6		Hierarchical (H)	FF-H	Medium	
7 8		Parallel (P)	FF-P		
9 10	IG (Information Gathering)	Hierarchical (H)	IG-H	Difficult	
11 12		Parallel (P)	IG-P		

Cognitive ability levels

Participants were tested for selected cognitive abilities. The tested abilities included operation span (a measure of working memory performance [4]. Their working memory abilities were split at median into two levels according to their scores: low (<60) and high (>=60).

3. RESULTS AND DISCUSSION

There were 48 participants in this experiment, and each of them searched for 6 search tasks, thus we have 288 searches in total. Among these 288 searches, participants issued only one query in 98 searches, which contained no query reformulations. The focus of the current paper is query reformulation in search sessions (N=190) that contained at least two queries, and how the frequency of individual reformulation types are related to task type, and users' cognitive styles.

The taxonomy of reformulation types

We constructed our own taxonomy by combining the types of query reformulation identified in prior work (e.g. [5], [7], [8], [10]) and based on users' queries in our study.

Table 2. A taxonomy of reformulation types

- Generalization (G)
  Q_i and Q_{i+1} contain at least one term in common; Q_{i+1} contains fewer terms than Q_i, and all terms in Q_{i+1} are shown in Q_i.
- Generalization with Reformulation (GR)
  Q_i and Q_{i+1} contain at least one term in common; Q_{i+1} contains fewer terms than Q_i, and Q_{i+1} also contains some terms that were not shown in Q_i.
- Specialization (S)
  Q_i and Q_{i+1} contain at least one term in common; Q_{i+1} contains more terms than Q_i, and also contains all terms in Q_i.
- Specialization with Reformulation (SR)
  Q_i and Q_{i+1} contain at least one term in common; Q_{i+1} contains more terms than Q_i, and also contains some terms that were not shown in Q_i.
- Word Substitution (WR)
  Q_i and Q_{i+1} contain at least one term in common; Q_{i+1} has the same length with Qi, but contains some terms that are not in Q_i.
- Repeat
  Q_i and Q_{i+1} contain exactly the same terms, but the format of these terms may be different
- Synonymous Reformulation (R2)
  Q_i and Q_{i+1} do not contain any common terms, but contains some synonymous terms.
- New (N)
  Q_i and Q_{i+1} do not contain any common terms or synonymous terms.

*Q_{i+1} is the following query next to the query Q_i in the same session.

We mainly considered the differences in the number of terms contained in two successive queries, as well as word substitutions. We found that these were the two most frequent reformulation types. In addition, we only considered the content change and we did not distinguish the format change. For example, if the user only changed the word order in a successive query, we classified it as the "Repeat" type. We also identified a new type of reformulation, which we believe had not been detected by other researchers – the "synonymous reformulation". The new type was identified by matching the terms in two queries according to WordNet.

Overall frequency of each reformulation type

Among the 190 search sessions, there were 495 reformulations in total. The overall frequency of each reformulation type is listed in Table 3.

Table 3. Overall frequency of each reformulation type

Reformulation types	Number of occurrence	Frequency
S	97	19.60%
WR	96	19.39%
G	75	15.15%
N	62	12.53%
SR	61	12.32%
GR	58	11.72%
R2	26	5.25%
REPEAT	20	4.04%

The above table shows that "Specialization" (S) was the most frequently used reformulation type, followed by "Word Substitution" (WR), and then by "Generalization" (G), "New" (N), "Specialization with reformulation" (SR), "Generalization with reformulation" (GR); while "Synonymous Reformulation" (R2) and "Repeat" were rarely used by users.

Frequency of query reformulation types by task type

To analyze the frequency of query reformulation types in different types of tasks, we first normalized the frequency of each reformulation type in each task session. For example, if there were 5 reformulations in one session, and 3 of them were "Specialization" (S), then the relative frequency of S in this

session was 0.6. Other frequencies were calculated in a similar way. Next we compared the mean frequency of each reformulation type between the different types of tasks.

A set of non-parametric Mann-Whitney U tests were conducted to examine the effects of task type features on the frequency of each reformulation type, because the distributions of the examined variables was not normal (Table 4).

Between "Fact Finding" and "Information Gathering" task types, "Synonymous Reformulation" (R2) was significantly different ($p<.05$). R2 was more frequently used in "Fact Finding" than in "Information Gathering".

Among "Simple", "Hierarchical" and "Parallel" tasks, two reformulation types showed to be statistically significant: "Specialization" (S) ($p<.05$) and "Word Substitution" (WR) ($p<.05$). In particular, "Specialization" (S) was more frequently used in "Simple" search tasks than in the other two types of tasks, and S was also more frequently used in "Hierarchical" tasks than in "Parallel" tasks. "Word Substitution" (WR) was more frequently used in "Parallel" tasks than in "Simple" tasks.

There was a significant effect of the task type on the frequency of "Specialization" (S) ($p<.05$), "Word Substitution" (WR) ($p<.05$) and "Synonymous Reformulation" (R2) ($p<.05$). In particular, S was more frequently used in FF-S and IG-H than in other tasks; WR was more frequently used in IG-P than in other tasks; R2 was more frequently used in FF-H than in other task types.

There was a significant effect of the objective difficulty level on two reformulation types: "Word Substitution" (WR) ($p<.05$), "Synonymous Reformulation" (R2) ($p<.05$). WR was much less frequently used in Easy tasks than in more difficult tasks, while R2 was more frequently used in Medium tasks than in either Easy or Difficult tasks.

There was a significant effect of the subjective difficulty levels on two reformulation types: "Generalization with Reformulation" (GR) ($p<.05$) and Repeat ($p<.05$). In particular, GR was more frequently used in Difficult tasks than less difficult tasks, while Repeat was more frequently used in Medium tasks than others.

In general, "Specialization" (S) was more frequently used in Simple tasks than in Parallel tasks. "Word Substitution" (WR) was more frequently used in Parallel tasks, especially Information Gathering with Parallel structure tasks, and tasks with high objective difficulty. "Synonymous Reformulation" (R2) was more frequently used in Fact Finding tasks with Hierarchical structure.

Frequency of query reformulation types by cognitive abilities

All the participants were categorized into two groups: people with higher cognitive abilities, and people with lower cognitive abilities. We compared the frequency of each reformulation type between these two groups using the non-parametric Kruskal-Wallis H test, but no statistical difference was found (Table 4).

4. FUTURE WORK

This study advances our understanding of how people reformulate queries when they conduct search tasks of different types. Our results suggested that the usage of some query reformulation types tended to vary between the features of task types, but did not vary between the searchers' individual differences. These findings will benefit research in query expansion and personalized search. We will examine the relationship between reformulation types and usefulness judgments, and consider other features of tasks and individual differences in the future work.

5. ACKNOWLEDGEMENT

This research was sponsored by IMLS grant LG#06-07-0105.

6. REFERENCES

[1] Aula, A. (2003). Query Formulation in Web Information Search. In Isaias, P. & Karmakar, N. (Eds.) Proc. IADIS International Conference WWW/Internet 2003, I, 403-410. IADIS Press. Retrieved on March 22, 2009, from: anne.aula.googlepages.com/questionnaire.pdf.

[2] Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, 5, 133-143.

[3] Ford, N., Miller, D., & Moss, N. (2005). Web search strategies and human individual differences: A combined analysis. Journal of the American Society for Information Science and Technology, 56(7), 757-764.

[4] Francis, G., Neath, I. (2003). CogLab on a CD. 3rd Ed. Wadsworth Publishing.

[5] He, D., Göker, A., and Harper, D.J. (2002). Combining evidence for automatic web session identification. Information Processing & Management, 38(5), 727-742.

[6] Hsieh-Yee, I. (1998). Search tactics of Web users in searching for texts, graphics, known items and subjects: A search simulation study. Reference Librarian, 60, 61-85.

[7] Huang, J. & Efthimiadis, E. N. (2009). Analyzing and Evaluating Query Reformulation Strategies in Web Search Logs. In Proceedings CIKM '09, 77-86.

[8] Jansen, B.J., Spink, A., Blakely, C., and Koshman, S. (2007).Defining a session on Web search engines. Journal of the American Society for Information Science and Technology, 58(6), 862-871.

[9] Larsen, B., Tombros, T. Malik, S. (2007). Interactive Track Guidelines for the Initiative for the Evaluation of XML Retrieval (INEX).

[10] Rieh, S. Y., & Xie, H. I. (2006). Analysis of multiple query reformulations on the web: The interactive information retrieval context. Information Processing and Management, 42(3), 751-768.

[11] Toms, E., MacKenzie, T., Jordan, C., O'Brien, H., Freund, L., Toze, S., Dawe, E., & MacNutt A., (2007). How task affects Information Search. In N. Fuhr, N. Lalmas, & A. Trotman (Eds.). Workshop Pre-proceedings in Initiative for the Evaluation of XML Retrieval (INEX) 2007, 337-341.

[12] Wildemuth, B. M. (2004). The Effects of Domain Knowledge on Search Tactic Formulation. Journal of American Society for Information Science and Technology, 55, 246-258.

Table 4. Mean frequency of each reformulation type by task facets and cognitive abilities

Reformulation types	Task Type		Task structure			Combined Task Type					Objective Difficulty			Subjective Difficulty			Cognitive Ability	
	FF	IG	S	H	P	FF-S	FF-H	FF-P	IG-H	IG-P	Easy	Med	Diff	Easy	Med	Diff	low	high
S	.21	.19	.31	.24	.12	.31	.19	.13	.26	.11	.31	.15	.19	.21	.18	.23	.19	.22
	z= .85, p=.36		**χ^2 = 10.10, p=.01**			**χ^2 =12.25, p=.02**					χ^2 = 3.61, p=.16			χ^2 = 1.72, p=.42			z= -.95, p=.34	
WR	.19	.28	.09	.20	.33	.09	.25	.24	.17	.42	.09	.26	.28	.25	.22	.20	.23	.23
	z= 1.91, p=.17		**χ^2 = 21.54, p<.01**			**χ^2 = 24.21, p<.01**					**χ^2 = 16.04, p<.01**			χ^2 = .37, p=.83			z= -25, p=.81	
G	.15	.09	.14	.12	.12	.14	.14	.16	.11	.07	.14	.16	.09	.12	.14	.13	.15	.10
	z= 3.54, p=.06		χ^2 = .59, p=.74			χ^2 = 5.75, p=.22					χ^2 = 3.74, p=.16			χ^2 = 3.13, p=.21			z= -.95, p=.34	
N	.14	.14	.11	.14	.16	.11	.14	.17	.13	.15	.11	.16	.14	.11	.14	.15	.15	.13
	z= .52 p=.47		χ^2 = 2.63, p=.27			χ^2 = 4.06, p=.40					χ^2 = 3.21, p=.20			χ^2 = .42, p=.81			z= .18 p=.86	
SR	.13	.12	.20	.11	.09	.20	.06	.09	.14	.09	.20	.08	.12	.11	.14	.12	.10	.15
	z= .05, p=.82		χ^2 = 1.51, p=.47			χ^2 = 6.58, p=.16					χ^2 = .18, p=.91			χ^2 = 4.87, p=.09			z= -1.58, p=.12	
GR	.08	.11	.10	.11	.08	.10	.04	.08	.14	.08	.10	.07	.11	.09	.08	.13	.11	.08
	z= 1.42, p=.23		χ^2 = .28, p=.87			χ^2 = 5.30, p=.489					χ^2 = 1.63, p=.44			**χ^2 = 6.37, p=.04**			z= .76, p=.44	
R2	.07	.03	.04	.08	.04	.04	.16	.06	.04	.01	.04	.09	.03	.05	.05	.08	.06	.04
	z= 4.28, p=.04		χ^2 = 2.11, p=.35			**χ^2 = 12.75, p=.01**					χ^2 = **11.22, p<.01**			χ^2 = .88, p=.64			z= -.22, p=.83	
REPEAT	.03	.04	.01	.02	.06	.01	.01	.05	.02	.06	.01	.04	.04	.03	.05	.01	.02	.04
	z= .00, p=.99		χ^2 = 5.08, p=.08			χ^2 = 6.21, p=.18					χ^2 = 2.93, p=.23			**χ^2 = 7.41, p=.03**			z= -.93, p=.35	

Personalizing Information Retrieval Using Task Stage and Task Type

Abstract:

Introduction and Related Work

Aimed at helping people find documents that meet their particular information needs, personalization of information retrieval (IR) takes account of information about users and their contexts beyond the queries that they submit. To avoid interrupting users, this additional information is often obtained implicitly from user behaviors and/or contextual data, such as dwell time, topic knowledge, and task information (Belkin, 2006). Dwell time, i.e., the time that a user spends on reading a retrieved document, has attracted much research attention. It was suggested that dwell time only is not a reliable factor for predicting document relevance in interactive IR, instead, it differs significantly according to specific tasks (Kelly & Belkin, 2004). White & Kelly (2006) found that information about tasks can be helpful in personalization, specifically, in setting a threshold for predicting web pages' relevance from dwell time.

Task stage has been found to affect user behaviors and search performance (e.g., Kulthau, 1991; Lin, 2001). Task type was also examined for its effect on user behaviors. One dimension of task type is task structure, along which are parallel vs. hierarchical tasks (in the study described in this paper, termed "dependent") (Toms et al., 2007). It is a pending issue whether or not contextual factors including task stage, task type, and topic knowledge can be helpful for implicitly predicting document usefulness, which is a key to personalization.

To this end, a longitudinal study was designed which aimed at answering the following research question (RQ):

RQ: Does the stage of the user's task help predict document usefulness from time spent on documents in the parallel and dependent tasks, respectively?

Methods

Twenty-four undergraduate journalism students were recruited as participants, each coming three times within a two-week period to a usability laboratory to do the experiment and get paid. They were asked to work on three sub-tasks associated with a single general task mimicking the journalists' assignment, couched either as parallel or dependent. Participants were asked to write a three-section article on hybrid cars, with each section to be finished at the end of each session. Half the participants worked on a dependent task, in which the accomplishment of some sub-tasks depends upon that of others. The three sub-tasks were: collect information on what manufacturers have hybrid cars, select three models that you will mainly focus on in this feature story, and compare the pros and cons of three models of hybrid cars. The other half participants worked on a parallel task in which the accomplishment of one sub-task does not depend upon that of others. The three sub-tasks were: finding information and write the report on Honda Civic hybrid, Nissan Altima hybrid, and Toyota Camry hybrid, respectively. In order to minimize the order effect, the study employed a careful design balancing the orders of participants' tasks as well as the orders of the three sub-tasks in the general task description.

In each session, participants were allowed to work up to 40 minutes to submit their reports. They can search freely on the Web to obtain the resources that they need for writing the reports. In order to ensure that the participants engage in a serious manner, they were told in the beginning that the top ¼ of whom have submitted the most detailed reports will receive a bonus.

Evaluation on the usefulness of each document was collected explicitly. In each session, after the participants submit the reports, they went through an evaluation process in which they assessed each document that they had viewed with regard to its usefulness to the overall task. They rated the usefulness based on a 7-point scale. Other data were collected by logging software Morae and pre- and post-session questionnaires. Morae records user-system interactions, including the time between each interaction events. For the purpose of ease to recording, the users were asked to keep only one Internet Explorer (IE 7.0) window open. They can use back and forward button to navigate between pages. The questionnaires elicit users' background information and perceptions on a number of aspects, such as their familiarity on the task topic.

Results

Total dwell time was used in the cases that a document was viewed several times over a session, which was the sum of the dwell time that a user spent on the document each time it was opened. In addition to total dwell time, there was another type of time examined in this study, which was called decision time. It was the time duration from when the document was opened till the point that the user made the first action, which usually indicated that the user had some decision on the usefulness of the page. For example, copying (and pasting) texts from a document usually meant that it was useful, while leaving a page without doing anything on it meant that it perhaps was not useful. For each RQ, both types of time, i.e., total dwell time and decision time were examined.

In general, data were analyzed in General Linear Model (GLM) since it takes care of the interaction effects among factors. Both types of time were transformed by logarithm using a base of 10 since their original distribution was far away from normal. For each RQ, data were analyzed generally in both tasks as well as in each task individually. Although the original usefulness score was collected based on 7-point scales from the users, they were collapsed into 3 groups because we think it would be enough from the system perspective to differentiate the within-variable differences by 3 groups, and the 7-point would be too fine.

Results indicated that when both tasks were considered, for total dwell time, usefulness was a significant factor ($p<.001$), but stage was not. However, for decision time, neither stage nor usefulness was a significant factor, but the interaction effect between stage and usefulness was significant ($p<.01$) contributing to decision time (Figure 1). In stage 1, users spent the shortest time for medium useful document and the longest time for very useful documents, but in stage 3, they spent the shortest time for very useful documents and the longest time for medium useful documents. This indicated that stage played some role contributing to decision time.

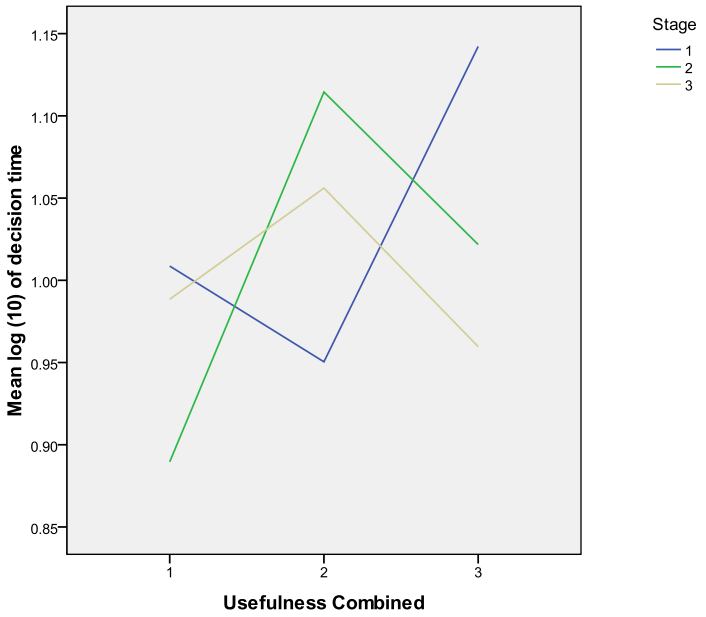


Figure 1. Document's mean log (10) of decision time for different usefulness score in different stages in both tasks

These relations were also examined in individual tasks. It was found that in the dependent tasks, usefulness was the single significant factors contributing to total dwell time ($p<.001$), as well as decision time ($p<.05$). Stage did not seem to play any significant role contributing to either type of time. Nevertheless, in the parallel task, the patterns were similar to those obtained when both tasks were examined. For total dwell time, usefulness showed significant effect ($p<.001$) but stage did not. For decision time, neither usefulness nor stage had significant effect, but their interaction had ($p<.05$) (Figure 2). In stage 1, users spent the shortest time for medium useful document and the longest time for very useful documents, but in stage 3, they spent the shortest time for very useful documents and the longest time for medium useful documents. Again, it indicated that in the parallel task, stage also played some role contributing to decision time.

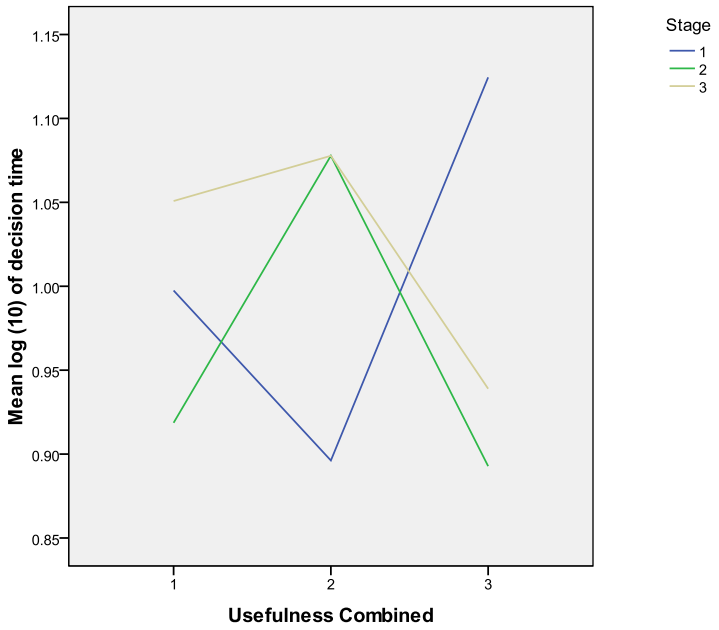


Figure 2. Document's mean log (10) of decision time for different usefulness score in different stages in the parallel task

Discussion & Conclusions

Through the examination of the relations among task stage, task type, document usefulness, and two types of time, it was found in this study that usefulness was the main factor contributing to total dwell time. This was reasonable considering the fact that the users often moved back and forth between reading the document and writing the report, and these documents which received longer dwell time were more likely to be useful. We also found that task stage plays some roles in contributing to decision time, especially in the parallel task and in both tasks in general. These findings are extremely helpful for personalizing search for specific users in that decision time can help predict the usefulness of documents given the task stage and task type information. In addition, unlike the total dwell time which would not be available until a session is over, decision time can be easily obtained since it is the time duration between opening the document till the user's first action, which can be easily captured by the system.

ACKNOWLEDGMENT

This research is sponsored by IMLS grant LG#06-07-0105.

REFERENCES

Belkin, N.J. (2006). Getting personal: Personalization of support for interaction with information. Talk in The 1st International Workshop on Adaptive Information Retrieval, October 2006, Glasgow, UK.

Chirita, P.-A., Firan, C.S., & Nejdl, W. (2006). Summarizing local context to personalize global web search. In Proceedings of CIKM '06, 287-296.

Kelly, D., & Belkin, N.J. (2004). Display time as implicit feedback: Understanding task effects. In *Proceedings of the 27th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '04)*, Sheffield, UK, 377-384.

Kuhlthau, C.C. (1991). Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42, 361-371.

Lin, S.-J. (2001). *Modeling and Supporting Multiple Information Seeking Episodes over the Web.* Unpublished dissertation. Rutgers University.

Teevan, J., Dumais, S.T., & Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. In Proceedings of SIGIR '05, 449-456.

Toms, E., MacKenzie, T., Jordan, C., O'Brien, H., Freund, L., Toze, S., Dawe, E., & MacNutt, A. (2007). How task affects information search. In N. Fuhr, N. Lalmas, & A. Trotman (Eds.). *Workshop Pre-proceedings in Initiative for the Evaluation of XML Retrieval (INEX) 2007*, 337-341.

White, R., & Kelly, D. (2006). A study of the effects of personalization and task information on implicit feedback performance. In Proceedings of CIKM '06, 297-306.

Community Interest Language Model for Ranking

Xiaozhong Liu
School of Information Studies
Syracuse University, Syracuse NY 13210
xliu12@syr.edu

Miao Chen
School of Information Studies
Syracuse University, Syracuse NY 13210
xliu12@syr.edu

ABSTRACT

Ranking documents in response to users' information needs is a challenging task, due, in part, to the dynamic nature of users' interests with respect to a query or similar queries. We hypothesize that the interests of a given user could be similar to the interests of the broader community of which she is a part at the given time and propose an innovative method that uses social media to characterize and model the interests of the community and use this dynamic characterization to improve future rankings. By generating community interest language model (CILM) for a given query, we use community interest to compute the ranking score of individual documents retrieved by the query. The CILM is based on a continuously updated set of recent (daily or past few hours) user-oriented text data while smoothed by historical community interest. The user-oriented data can be user blogs or user generated textual data.

General Terms

Algorithms, Human Factors, Experimentation

Keywords

Information Retrieval, Ranking, Topic, User, Blog, Community Interest, LDA, Ranking

1. INTRODUCTION

Ranking is a key step in Information Retrieval (IR) systems. Existing ranking algorithms use different approaches to increase performance based on similarity computation, social link analysis, user behavior data, or personalization (user profiling). Ranking is a dynamic problem, namely, user judgments with respect to a query may change dramatically over time. We hypothesize that the ranking score for each retrieved document in the search result should depend on current community interests, for instance, as the following formula shows, (ranked by) the probability that the community (for the target query) interested in document at a given time.

$$Score(doc) = P_{interest}(doc|Community_{query}, time)$$

In this paper, we use "*community interest*" to determine the ranking score, and we compute the interest level of the global (or local) community in a specific document for a given query at a given time. Instead of employing user judgments about what is interesting and what is not, we will use *user oriented (real-time) text data* (such as blog postings, news comments or user selected news text) to represent users' interests. By using a topic-modeling algorithm, *topics* of the real-time community interest in the user text data are identified as probability distributions over words. Each word or topic is then weighted by historical text data from the community. At last, the community interest language model (CILM) is constructed as a language model for each query to represent the current interests of the community. For each document in the search results, we also infer a score (using the precomputed probability topic models) that is proportional to the level of community interest in this specific document given the query. This score is then used for ranking the entire set of retrieved results.

2. COMMUNITY INTEREST RANKING

In the Web 2.0 context, users may generate different kinds of text data, such as blogs, selected news, and comments that reflect their interests. In this paper, we use time sensitive blog data (from blog search engine) to represent users, and we also extract dynamic computational community interest from language model perspective.

For each popular query (from query log), a list of real-time blogs is collected. Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) is used to extract the topics within the collection, and each topic is a probability distribution over words. The next step is to model the community interest based on the extracted topics for ranking.

2.1 Community Interest Language Model

We define community interest (toward each query) as a dynamic probability distribution of each candidate topic over each query, and each number in this distribution represents the current community interest probability of a specific topic given the target query. From language model perspective, the final ranking score could be the (retrieved) document likelihood given the dynamic topic probability distribution for the target query.

$$Score(doc) = P_{interest}(doc|\theta_{query,topic-interest})$$

As mentioned earlier, the candidate topics are extracted from most recent blogs generated by users (in the community). In the preliminary experiment, we find there are three different kinds of topics:

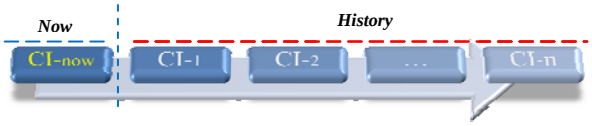
1. *Background topic (stoptopic): the topic covers the very basic background features of the query. Those words could be judged as a query specific stopword list.*
2. *Hot topic: there are two types of hot topics for the community; first, a topic in which the community is continuously and*

increasingly focused, and second, a topic related to breaking news surrounding the query, which is of great interest in the community.

3. *Diminishing topic: the topic is no longer popular for community; and the community's interest is shifting to other topic(s).*

In order to detect each abovementioned topic category and build language model to mirror the current real-world community interest, we will use historical community interest as the smoothing factor.

First, the collected user generated blog data will be separated into different segments based on their generate timestamp (as the following diagram shows, from *CI-now* to *CI-n* ordered by time). Each segment can be viewed as a snapshot of the community interest (about the query) at a given time. Second, the most recent LDA topic model will be used to infer each of the historical CI to get the probability of each CI-k interested in topic-k, *P (topic-z | CI-k).*



In this paper, from language model perspective, retrieved document likelihood given real-time community interest distribution θ_{CI} estimated by real-time blogs will be used to rank the result collection. As the following formula shows:

$$Score(doc) = P(doc \mid \theta_{CI})$$

CI is the real time community interest

As mentioned earlier, two different parts compose community interest (*CI*): *CI-now* (the most current *CI* snapshot) and *CI-history* (historical *CI* snapshots). According to (Zhai & Lafferty, 2004), smoothing could be used to model the noisy word (like IDF effects) and improve the accuracy of the estimated language model. In this research, we face the same problem, as we need to model the real-time community interest by identifying the popular topic distribution, while filtering the noisy information (background and diminishing topics). In order to solve this problem, we used history community interest distribution as the smoothing factor as following:

$$\log P(doc|\theta_{CI}) = \log \frac{P(doc|\theta_{CI-now})}{\alpha \cdot P(doc|\theta_{CI-history})} + n \cdot \log\alpha$$

In the above formula, the probability that *CI* generate the document could be computed by *CI-now* generate document probability divided by *CI-history* generate document probability. So, if a document gets a high ranking score, it should have a high current interest probability score (document topics interest current community) with a low historical interest probability score (document topics not interest historical community).

As we know, *CI-now* is the most current LDA model based topic distribution. So, the problem left is how to model *CI-history*. Intuitively, historical community interest should be a decay function, as the more recent community snapshots should have a larger contribution to the interest model compared with old snapshots. Based on this hypothesis, the following recursive function will be employed to define to *CI-history* with user interest decay parameter β.

$$\theta_{CI-n} = \theta_{n-topic}\ ;\ \theta_{CI-k} = \frac{\theta_{k-topic} + \beta \cdot \theta_{CI-(k-1)}}{1+\beta}$$

In the formulas, θ_{CI-n} is the oldest community interest snapshot, and it is estimated by the inferred topic probability distribution. For any θ_{CI-k}, it is defined (normalized) by θ_{n-1} and CI_n with an interest decay parameter β. Based on this definition, finally, the *CI-history* will be:

$$\theta_{History} = \theta_{CI-1} = \frac{\theta_{1-topic} + \beta \cdot \theta_{CI-2}}{1+\beta}$$

So, CI_1 is composed by $CI_1, CI_2, CI_3 \ldots CI_n$ (all the historical *CI* snapshots) and decay parameters will be $1, \beta^2, \beta^3 \ldots \beta^{n-1}$ ($0 < \beta < 1$).

Finally, the retrieved documents will be ranking by the real time community interest generate probability scores.

From July to December 2008, I implemented the CIV approach with blog training at Yahoo! Search. And in the next academic year, I will be focusing on CILM approach development and experiment with blog training as well as comment tagged news training.

3. EVALUTION

The evaluation of a ranking algorithm is difficult, especially for the real-time ranking task, which cannot employ existing test collections such as TREC. Precision-at-document-n (Anh & Moffat, 2002) is currently a good measure for the web, as most users will be focusing on only the very first page of n results. And Normalized Discount Cumulative Gain (NDCG) (Järvelin & Kekäläinen, 2002) works when user graded relevance data is available.

As a real-time ranking algorithm proposed in this thesis, human judgment will be used for evaluation. The CIV and CILM generated from blog and news will be compared to the popular ranking algorithms and existing web search results. The top ranked documents will be rated as "interesting", "just ok", or "not relevant" by user. For NDCG, each kind of user judgment will be assigned a numeric score, such as 2, 1 and 0. In preliminary evaluation, experiment with Yahoo collection, the CIV algorithm can increase the number of "interesting" ratings by 16.74% while decreasing the "not relevant" rating by 20.59% compared with popular news search engine ranking result (over 9 queries, top 5 search results by 5 users for 5 days). NDCG shows that CIV can significantly (t-test $p<0.05$) increase the ranking performance.

In the next academic year, I will launch a much more comprehensive user evaluation based on Amazon Turk and propose an innovative dynamic ranking evaluation method. While traditional evaluations are based on static relevance judgments, the new ranking evaluation will be based on a user's real-time preferences. As time is introduced as a new parameter in the evaluation matrix, a user's judgment about the same query-document pair could be different depending on the moment the decision is made. The great challenge is to develop an evaluation framework that allows results from the new ranking algorithms to be compared to those obtained using existing ranking algorithms and to popular search engine ranking results.

4. REFERENCES

[1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. The Journal of Machine Learning Research, 3.

[2] Liu, X., & Brzeski, V. v. (2009). Computational community interest for ranking. Paper presented at the Conference on Information and Knowledge Managemen

[3] Anh, V. N., & Moffat, A. (2002). Improved retrieval effectiveness through impact transformation. Paper presented at the ACM International Conference Proceeding Series.

[4] Järvelin K. & Kekäläinen J., Cumulated gain-based evaluation of IR techniques, ACM Transactions on Information Systems (TOIS), v.20 n.4, p.422-446, October 2002

[5] Zhai, C., & Lafferty, J. (2004). A Study of Smoothing Methods for Language Models Applied to Information Retrieval. ACM Transactions on Information Systems (TOIS), 22(2), 179-214.

Wayfinding in the Labyrinthine Library: A Mixed Methods Study Investigating Public Library User Wayfinding Behavior

Lauren H. Mandel
Florida State University
142 Collegiate Loop, Po Box 3062100
Tallahassee, FL 32306-2100
LMandel@fsu.edu

ABSTRACT

The design of public library facilities is an important area of concern for public librarianship. The building is the physical expression of the library's mission and purpose, to provide information and services to users. If users cannot access and utilize the facility effectively, then they also cannot access and utilize the library's resources and services. The large body of literature devoted to public library facility design shows the importance the field places on this issue.

Wayfinding is the method by which humans orient and navigate in space, and particularly in built environments such as cities and complex buildings, such as public libraries. In order to wayfind successfully in the built environment, humans need information provided by wayfinding systems and tools, such as architectural cues, signs, and maps. This is true of all built environments, including public libraries, but the issue is all the more important in public libraries where users already enter with information needs and possibly anxiety, interfering with the ability to wayfind successfully. To facilitate user wayfinding, which in turn facilitates user information-seeking, public library facilities need to be designed with users' wayfinding needs in mind, along with consideration of users' information-seeking and other library-specific needs.

From his research, Passini observed that the decision plan is developed according to five problem-specific strategies and utilizing two user-specific styles that he detailed in his Conceptual Framework of Wayfinding, the theory guiding the proposed dissertation reported in this poster [1]. The strategies correspond to information seeking (and other problem-solution) strategies and are more or less observable, depending on the strategy, method employed for data collection, and level of forthcoming of the research subjects. The same is true of the two wayfinding styles; they correspond to information seeking styles and have varying degrees to which they may be observed.

In order to research the problem-specific strategies and user-specific styles in the context of public library users having to wayfind while information-seeking, each must be made observable and measurable. Passini does not specify research methods to employ while observing each of these strategies and styles.

Passini sees the wayfinding decision plan as a structured process that operates at different levels of generality, through which the wayfinder focuses on individual tasks or subtasks always while considering the problem as a whole (Strategy 1: Dividing the Task into Manageable Parts While Keeping an Eye on the Larger Task at Hand). But, he can only deal with one problem or subtask at a time (Strategy 2: Narrowing), following a continuous process that can deal with unforeseen problems whenever they occur, pointing to the dynamic property of decision making (Strategy 3: Adapting and Responding). For as large a part of the decision plan as possible, the wayfinder relies on an existing solution repertoire (Strategy 4: Accessing One's Schemata). He also bases his plan on the available environmental information (Strategy 5: Gathering Information and Adapting Accordingly).

The public library facility design literature identifies the importance of understanding patron wayfinding behavior and designing around it, and the purpose of this proposed dissertation is to make a step toward answering that call. A single-method pilot study utilized unobtrusive observation to investigate library users' initial wayfinding behavior from the two entrances of a medium-sized public library, with the data analyzed and displayed using Geographic Information Systems (GIS) software. The pilot study found that certain routes are more popular than others and suggested that such information can be gathered relatively easily and then used by the library to improve the library's wayfinding system and for marketing of library materials in high-traffic areas.

However, the pilot study's largest limitation, namely the inability to ascertain any user opinions regarding their wayfinding in the library, indicates the need for the proposed dissertation to employ a mixed method research design that replicates the original unobtrusive observation and adds in-depth interviews. This will allow the dissertation research to offer a more comprehensive view of library user wayfinding

behavior, particularly understanding of how users implement Passini's wayfinding strategies while orienting and navigating in public libraries. The overall purpose of this mixed method research is to explore how users navigate from the entrance of a library, which routes are most popular and areas that experience the highest traffic, what methods users employ to conduct this navigation, how users feel about their ability to wayfind (or not) in the facility, and ways they would like the existing wayfinding system to be altered (if any). The goal is to explore these topics as a beginning to developing a redesign plan for the library serving as the research site (henceforth, the Library) that improves the facility's ease of wayfinding and overall usability and provides suggestions for library marketing efforts that make use of the knowledge of the most popular routes and highest trafficked areas of the library.

The poster details the dissertation prospectus, which proposes a mixed method research design, guided by Passini's Conceptual Framework of Wayfinding, to investigate library user wayfinding behavior from the entrance of a medium-sized public library facility. The mixed method design includes two methods, unobtrusive observation of library user wayfinding behavior and in-depth interviews with library users to discuss their views on wayfinding in public libraries. See Figure 1 for a graphic depiction of the mixed method research design guiding the dissertation. A mixed method design is chosen to guide this dissertation because of the ability to triangulate data gathered from different methods, thereby mitigating the limitations of a single-method dissertation, strengthening the findings, and providing a more comprehensive view of library user wayfinding behavior than could be obtained from a single-method approach.

Categories and Subject Descriptors

H.1.m [**Information Systems**]: Models and Principles – *miscellaneous.*

General Terms

Design, Human Factors.

Keywords

Wayfinding, public libraries, library facilities.

REFERENCES

[1] Passini, R. 1981. Wayfinding: A conceptual framework. Urban Ecology, 5 (1981), 17-31.

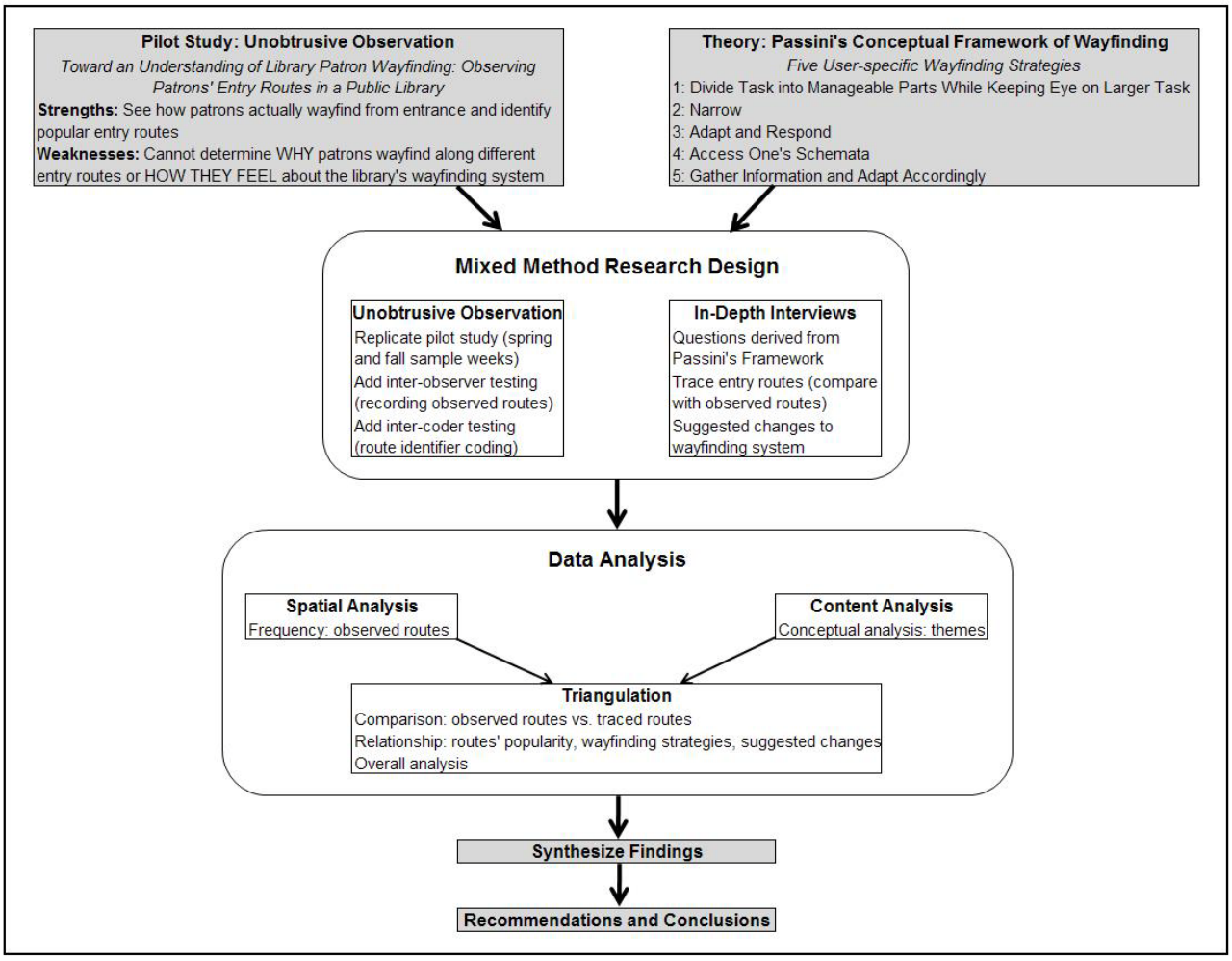


Figure 1. Graphic depiction of the mixed method research design.

Using Maps for Virtual Collaboration: Implementing a Framework for Visually Enabled Geocollaboration

Janet Marsden
iSchool
Syracuse University
337 Hinds Hall
315-443-5509
jamarsde@syr.edu

ABSTRACT

The purpose of this study is to test whether the use of graphic information interfaces such as maps and simulations can enhance the general public's understanding of scientific data and risk analysis, improve communication and virtual collaboration, and lead to environmental decision making that is less positional and more consensual. The research will be located in upstate New York in the Adirondack Park region, and will focus on land use proposals such as cellular telephone tower installations within and near the Park's Blue Line that have been issues of contention. This design-based research study proposes to investigate the use of geospatial and telecommunications technologies (web-deployed maps and simulations) for presenting scientific data and analyses in a way that balances stakeholders' interests and promotes understanding. The research addresses the need for empirically-based guidelines for presenting the results of scientific inquiry to the public to support rational decision-making. The investigation will result in more effective virtual collaboration between civic leaders, planners and communities, leading to better ways to engage stakeholders on a broad range of environmental and urban planning, resource management, and education issues through leveraging geospatial and telecommunications technologies.

1. INTRODUCTION

The purposes of this study are to develop guidelines for using geospatial technologies such as digital maps, simulations and graphic scripts to present scientific data and analyses in a way that promotes rational decision-making and balances stakeholders' interests; and to improve communication and collaboration between business, government, scientific communities and the public on issues related to scientific inquiry, risk analysis, environmental and community planning, resource management, and education by leveraging geospatial and telecommunications technologies. In addition, this research will further practices of collaborative governance, collaborative problem solving, and collaborative public management. From a pragmatic standpoint, it's expected that this approach will elevate public discourse, reduce litigation and legal actions and contribute to positive conflict outcomes. Finally, this investigation will complement research findings in the field of education that suggest the use of maps, simulations and information visualizations can lead to improved learning outcomes.

2. BACKGROUND

In any problem involving multiple stakeholders there seems to be inevitable conflict. Two ways that ICT can promote understanding and rational decision-making are: improving access to information through telecommunications technologies, and improving the quality of the information made available. This project is an investigation into the use of web-deployed map-based interfaces as a contextualizing medium for virtual collaboration. Questions include: Can geospatial technologies broaden public discourse about land use issues and help resolve conflicts? Can geospatial technologies such as digital maps and simulations help participants better understand scientific data and risk analysis, make better informed decisions, and avoid costly litigation?

MacEachren and Brewer call for research into the role of geovisualization as an enabler for needed advances in virtual collaboration "… developments in geographic information science, and in computer graphics and visualization, suggest that we are [also] on the cusp of a substantial increase in the role of maps, images, and computer graphics as mediators of collaboration, in a range of contexts including scientific inquiry, environmental and urban planning, resource management, and education." Palmer and Smardon (1989) discuss the shortcomings of traditional public participation in environmental management. One of these is the lack of timely participation by the majority of citizens. Their research also showed that those who do attend public meetings tend to be more activist and hold more extreme positions than the general public, making issues appear to be more contentious than they are perceived by the majority. Some use misinformation and scare tactics to sway the opinions of others. Without reliable information from trusted sources, people may tend to make decisions that are not based on facts. The reasons that many people don't attend are numerous and include time, transportation, childcare and other resource constraints. Morgan, et al point out that people usually don't become concerned about issues until they perceive a potential threat or risk to themselves or others that they feel connected to. When, after the fact, they become aware of an impact or perceived risk to themselves or others and react, their position is frequently defensive, and based on fear rather than rational thinking. Simply conducting studies and producing research results doesn't mean that the general public is aware of them, can find them, or that they can correctly interpret and understand them when they are available. It's unlikely that when trying to make a decision about leasing mineral rights or using pesticides on their lawn that people will search for and read research articles published in peer-reviewed journals, analyze data or pore over complex formulae. But 68% of people between the ages of 50 to 64, and higher percentages of those under 50, turn to the Internet to get information when they want to

answer a question about health, the environment or other issues (AARP, December 2009).

3. CONCEPTUAL FRAMEWORK

MacEachren and Brewer (2004) define a conceptual framework for studying "geocollaboration", defined as "visually-enabled collaboration using geospatial information through geospatial technologies". The three elements of the framework are visualization of data and information, virtual collaboration and the use of map-based interfaces for virtual collaboration.

4. VISUALIZATION

The research purpose is to test geospatial technologies such as GIS, GPS and interactive maps as effective visual media for virtual collaboration. This research is important as we become a more global society and as this leads us to collaborate locally, regionally, nationally and internationally. Ironically, as MacEachren and others call for research on geocollaboration, Jerome Dobson (2007) points out the abysmal level of geographic knowledge possessed by a generation that has not learned to use maps (Dobson, 2007), the lack of geography lessons in K-12 education, and the ongoing loss of geography departments in US colleges and universities as the demand for these skills continues to increase; and calls for us to 'Bring back Geography!'. Of the vast quantities of digital data being generated today, it has been estimated that as much as 80% of it includes geospatial referencing such as geographic coordinates, addresses or postal codes (MacEachren and Kraak, 2001). Representing this data in a meaningful way for its use in problem-solving, education, and conflict resolution presents a challenge. "Computational and experimental sciences produce and collect ever-larger and complex datasets, often in large-scale, multi-institution projects. The inability to gain insight into complex scientific phenomena using current software tools is a bottleneck facing virtually all endeavors of science" (Aragon et al, 2008). This is an even greater challenge in the effort to enable non-scientists to understand the complexities of scientific phenomena. Although research has shown that data visualization technologies can facilitate comprehension of complex scientific results (Aragon, et al., Dunleavy, et al.), more work is needed.

5. VIRTUAL COLLABORATION

Veinott, Olson, and Fu investigated how virtual teams benefit from video and audio technology vs. audio only in solving problems. Their research showed that visual cues (i.e. seeing each others' faces) enhance understanding, coordination and teamwork for non-native English speakers trying to work together on fairly complex problem-solving tasks. The following quote is from the abstract: "We compared the performance and communication of people explaining a map route to each other. Half the pairs have video and audio connections, half only audio. Half of the pairs were native speakers of English; the other half were non-native speakers (who presumably would have to negotiate meaning more). The results showed that non-native speaker pairs did benefit from the video; native speakers did not. Detailed analysis of the conversational strategies showed that with video, the non-native speaker pairs spent proportionately more effort negotiating common ground." In their study, the authors gave maps to each pair that were slightly different, but the pairs could not see each other's maps. Veinott et al investigated whether seeing facial expressions could enhance language understanding, but found no benefit except for non-native English speakers. What would the outcome have been if the pairs were tested on their use of the maps they were given as visual cues rather than the faces of their partners? As we look at examples such as the use of email, social networking, and the Internet, there is no doubt that modern telecommunications applications have revolutionized the nature of collaboration. Research has shown that virtual collaboration is regarded as a mixed blessing. When it allows participation, communication and collaboration where it would otherwise be impossible, it's seen as a benefit; but many people still prefer face-to-face meetings (Beyond Being There, 2008). However, as more and more people become accustomed to using the Internet and telecommunications, these media are increasingly accepted as alternatives to traditional meeting venues.

6. MAPS FOR COLLABORATION

Couclelis and Monmonier (1995) discuss using SUSS, a map-based information system designed to be used as an analysis and communication tool for resolving contentious land use issues. Graphic scripts are visual narratives that use sequenced map displays to present data dynamically (Monmonier, 1996). Interactive scripts allow users to query, zoom, pan or review. The authors define SUSS as a problem structuring system rather than a decision-making (or decision-support) system, i.e. it is meant to be used for understanding complex or contentious problems from multiple perspectives. They propose using a GIS-based system with a political negotiation metaphor as its organizing principle. This study advances Couclelis and Monmonier's approach to developing map-based interfaces and applies them to test the role of geovisualization in contextualizing problem-based scenarios for virtual collaboration. Risk communication with the public about environmental policy and decision making is often misunderstood, distrusted or rejected because of the use of jargon, probabilistic conditioning of conclusions based on legitimate scientific uncertainties or unexplained references to opposing arguments or positions (Morgan et al, 2002). Decision making that is less positional and more consensual can be achieved by addressing trust issues through the use of tested and verified data and information with thorough documentation (Ury et al, 1991, O'Leary et al, 2003). Based on studies conducted on the use of simulations for teaching and learning (Hakkareinen, 2007, Dunleavy et al, 2008), both teachers and students have reported that the use of information visualizations and simulations in a collaborative problem-solving situation can afford a highly engaging, collaborative learning experience that is compatible with multiple information-seeking and problem-solving styles. The importance of visual cues for achieving cooperation and agreement in collaborative actions, both virtual and face-to-face, is well-documented (Kraut et al, 2003, Veinott et al, 2009).

7. METHODOLOGY

A design-based research approach is planned. Cobb et al (2003) describe design-based research as "pragmatic as well as theoretical in orientation – in that the study of function – both of the design and of the resulting ecology of learning – is at the heart of the methodology." It is an empirical technique which involves designing interventions with specific goals or objectives, testing them, evaluating the results, then refining or adjusting the intervention. It is particularly appropriate for this study because it

is cognizant of the challenges of multi- and inter-disciplinarity, and because it recognizes the value and importance of context for problem-solving. The design strategy for this study has two components. The first involves designing a problem-based scenario for siting cell phone towers. The second is designing a map-based graphic narrative for presenting the scenario. Graphic scripts will be deployed with the problem-solving scenario in a virtual environment. Participants will use the interactive narrative for virtual collaboration to find out if the use of maps as visual cues can improve virtual collaboration by providing context through the visual presentation of relevant information. Content will be developed based on Monmonier's guidance (Monmonier, 1993, 1996).

For this study, approximately 160 (40 groups of two to four) participants will be recruited from various stakeholder groups concerned with proposals for land use in locations within and around the Adirondack Park Blue Line. Participants will work in randomly assigned groups of two to four people. The sessions will be virtual collaborations rather than face-to-face. Participants will be instructed to work in their groups to understand the information provided, answer questions about the issue, and report on any agreement reached. All participants will have access to the online maps and information and will be able to communicate with each other, but will not be co-located and will not be able to see each other. The graphic interface will serve as the visual context for problem-solving within each group, and each group will negotiate discussion among themselves. Sessions will be recorded to capture audio, and screen capture will be used to record text-based communications, time frames, navigation and group actions. Participants will be asked to complete an initial short questionnaire to collect demographic data such as gender, age, race, education, etc., their position on or interest in the issue, and self-assessments of their experience with and understanding of maps prior to the session. The sessions will be designed to last around one and a half hours in all. Participants will receive a small incentive ($20 value) upon completion.

The sessions will be evaluated on the basis of quantitative and qualitative measures of the participants' use of the information provided, and the outcomes of the collaborations. Participants will be asked to complete a short follow-up questionnaire about their perceptions of the process and its outcome. Selected participants may be asked for follow-up interviews, or may attend a debriefing meeting if they wish.

8. ACKNOWLEDGMENTS

My sincere thanks to Dr. Ruth Small, Dr. Mark Monmonier, Dr. Derrick Cogburn, Dr. Steven Sawyer and Dr. Alan Foley for their essential help and support.

9. REFERENCES

[1] AARP Bulletin, December 2009, Vol. 50, No. 10, pp. 4.

[2] Aragon, C., Bailey, S., Poon, S., Runge, K. and Thomas, R. "Sunfall: a collaborative visual analytics system for astrophysics", Proceedings of SciDAC, 2008.

[3] "Beyond being there", Final Report from the Workshop on Building Effective Virtual Organizations, May 2008.

[4] Cobb, P., Confrey, J., diSessa, A., Lehrer, R., Schauble, L. "Design Experiments in Educational Research", Educational Researcher, Jan-Feb 2003, Vol. 32, No. 1, pp. 9-13.

[5] Couclelis and Monmonier, "Using SUSS to Resolve NIMBY: How Spatial Understanding Support Systems Can Help With The "Not In My Back Yard" Syndrome", Geographical Systems, 1995, Vol. 2, pp. 83-101.

[6] Design-based Research Collective. "Design-based Research: An Emerging Paradigm for Educational Inquiry", Educational Researcher, Vol.32, No. 1, pp. 5-8.

[7] Dobson, J. "Bring Back Geography!", ArcNews Online, Spring 2007, ESRI.

[8] Dunleavy, M., Dede, C. and Mitchell, R. "Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning", Springer Science+Business Media LLC, 2008.

[9] Garrard, S., Turner, K. and Bateman, I. Environmental Risk Planning and Management, Edward Elgar Publishing, LTD, Cheltenham, UK, 2001.

[10] Gersmehl, P. Teaching Geography, 2nd edition, Guilford Press, New York, NY, 2008.

[11] Gersmehl, P. and Gersmehl, C. "Wanted: A Concise List of Neurologically Defensible and Assessable Spatial Thinking Skills", Research in Geographic Education 8:5-38, 2006.

[12] Hakkareinen, P. "Designing and Implementing a Problem-Based Learning course on educational digital video production: Lessons learned from a design-based research". Association for Educational Communications and Technology, 2007.

[13] Jacobson, D. "Doing Research in Cyberspace", Field Methods, Vol. 11, No. 2, Nov. 1999, pp. 127-145.

[14] Kraak, J. and Ormeling, F. Cartography: Visualization of Geospatial Data, 2nd edition. Pearson Education Prentice Hall, Harlow, England, 2003.

[15] Kraut R., Fussell S., and Siegel J. "Visual Information as a Conversational Resource in Collaborative Physical Tasks".

[16] MacEachren, A. and Brewer, I. "Developing a conceptual framework for visually-enabled geocollaboration", Int. J. of GIS, Vol. 18, No. 1, 2004, pp. 1-34.

[17] MacEachren, A. and Kraak, J. "Research challenges in geovisualization", Cartography and GIS, Vol. 28, No. 1, 2001, pp. 1-11.

[18] Monmonier, M. How to Lie With Maps, 2nd edition, University of Chicago Press, Chicago, IL, 1996.

Monmonier, M.

[19] Mapping It Out: Expository Cartography for the Humanities and Social Sciences, University of Chicago Press, Chicago, IL, 1993.

[20] Morgan, M., Bostrom, B., Fischman, A. and Atman, C. Risk Communication: A Mental Models Approach, Cambridge University Press, 2002.

[21] O'Leary, R. and Bingham, L. The Promise and Performance of Environmental Conflict Resolution, Resources for the Future Press, 2003.

[22] Palmer, J. and Smardon, R. “Measuring Human Values Associated with Wetlands: Comparing Public Meetings and Sample Surveys”, in Intractable Conflicts and Their Transformation, edited by Louis Kriesberg, Syracuse University Press, 1989.
[23] Tufte, E. “The Visual Display of Quantitative Information”, Ch. 4, Graphics Press, Cheshire, CT.
[24] Ury, W., Fisher, R. and Patton, B. Getting to Yes, 2nd edition, Houghton-Mifflin, New York, NY, 1991.
[25] Veinott, B., Olson, J.S., Olson, G.M., & Fu, X. “Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other”, Proceedings of CHI 1999, pp. 302-309.

Assessing iSchool Effectiveness through Alumni Feedback: Preliminary Results from the Workforce Issues in Library and Information Science 2 (WILIS 2) Project

Joanne Gard Marshall (1) Jennifer Craft Morgan (2); Victor W. Marshall (2); Deborah Barreau (1); Barbara Moran (1); Paul Solomon (3); Susan R. Rathbun-Grubb (2); Cheryl A. Thompson (2)
Organization(s): 1: School of Information and Library Science, University of North Carolina, United States of America; 2: Institute on Aging, University of North Carolina, United States of America; 3: School of Library and Information Science, University of South Carolina, United States of America

Workforce Issues in Library and Information Science 2 (WILIS 2) is an IMLS funded project designed to implement a career tracking model for Library and Information Science (LIS) graduates. This project was introduced to the iSchool community via a poster session at the 2009 iConference and recruited the participation of **8 iSchools** and 36 other LIS programs in the US and Canada to test the career tracking model and program evaluation instrument. Each program was funded to survey up to 250 of its recent graduates. Data are currently being collected and the response rate so far is 45.4%. The aggregated preliminary results will be presented in this poster, and will showcase characteristics of recent graduates and evidence of iSchool program effectiveness. Career results will include such items as length of job search, current job settings and titles, job satisfaction, leadership responsibilities, professional contribution, and continuing education needs. Program evaluation results will include respondents' overall satisfaction with their program, sense of preparedness for the workplace, and suggestions for LIS program improvement.

Project Background: LIS programs have generally lacked the time and resources to systematically survey their graduates; as a result, stakeholders lack an adequate understanding of what happens to their students after they graduate. Educators, in particular, do not have ongoing data about the extent to which their programs meet students' expectations, prepare them for the workplace or meet continuing learning needs. Such an understanding will assist in educating and managing the LIS workforce more effectively.

WILIS 2 builds on WILIS 1, a comprehensive study of career patterns of graduates of LIS programs in North Carolina. Using a Community Based Participatory Research (CBPR) approach, WILIS 2 refined the WILIS 1 career tracking model so that it would be suitable for use by any LIS program. The WILIS 2 implementation has provided an opportunity to experiment with the use of CBPR to create a national career tracking model that will have the best chance of being widely adopted and used by LIS programs. Evidence-based workforce and educational planning are both essential to supporting the successful recruitment, education and retention of the next generation of LIS professionals. The stakeholders who require the information necessary to do such planning include LIS educators, professional associations, LIS employers, practicing professionals, LIS students and the larger library and information community. Each of these groups has somewhat different information needs and interests in the process of workforce and educational planning. Recognizing the importance and perspective of these differences, the study team has utilized elements of CBPR to maximize consensus and buy-in. This process has resulted in the creation of a survey that will create an evidence base from which stakeholders can

make important decisions related to the recruitment, education, reentry and retention of future LIS professionals.

CBPR is "a collaborative approach to research that equitably involves all partners in the research process and recognizes the unique strengths that each brings. CBPR begins with a research topic of importance to the community and has the aim of combining knowledge with action and achieving social change (Isrel, et. al., 1998)." While CBPR is largely used in public health research, it makes sense for WILIS 2 as a way to bridge the education and professional practice worlds of LIS in constructive and mutually beneficial ways. For WILIS 2, the study team has worked side by side with the stakeholders to do all of the following: 1) craft the WILIS 2 survey; 2) finalize methodological tools; 3) implement the launch of the WILIS 2 model; 4) disseminate the findings; and 5) explore sustainability options. A likely side product of this work will be to increase the capacity of the stakeholder partnership created through the formation and experience of a Project Advisory Committee (PAC) to work together on other important research and practice issues in the field. This approach also holds promise for narrowing the gap between researchers, educators and practitioners.

The specific research goals of WILIS 2 are to:
1) Refine a career tracking model that is suitable for all LIS programs to use with their recent graduates. Using the CBPR approach and a PAC of 16 members including LIS program directors, library leaders and workforce experts, we have built consensus on the essential elements of a career tracking instrument for use in monitoring and planning both educational programs and LIS workforce needs.

2) Recruit as many LIS programs as possible to participate in a phased national launch of the career tracking model. The study team, in concert with the PAC, recruited LIS programs in the US and Canada to participate in an initial survey launch phase of 8 programs, a second phase of 20 programs, and a third phase of 16 programs.

3) Conduct the surveys and provide access to results for the participating LIS programs. Participating LIS programs will have access to their own datasets, automated reports, analytic tools and analytic benchmarking feedback from the study team.

4) Explore options for sustaining the national career tracking model. The PAC, in conjunction with the study team, will direct and conduct the exploration of avenues for sustainability beyond the grant period. This process will result in a final plan for sustainability generated by the PAC and the study team and systematically presented to other LIS stakeholders not involved in the creation of the plan.

5) Disseminate findings and publicize the availability of the WILIS 2 model. The study team members and the PAC will collaborate to explore avenues for dissemination of this research, such as conferences and a Webcast.

Works Cited

Israel, B.A., Schulz, A.J., Parker E., & Becker, A.B. (1998) in "Review of Community-Based Research: Assessing Partnership Approaches to Improve Public Health, *Ann. Rev. Public Health,* 19:173-202.

Orality in the Library: How Mobile Phones Challenge Our Understandings of Collaboration in Hybridized Information Spaces

Rhonda McEwen
University of Toronto
Faculty of Information

Kathleen Scheaffer
University of Toronto
Faculty of Information

rhonda.mcewen@utoronto.ca kathleen.scheaffer@utoronto.ca

ABSTRACT

Following a period of rapid and widespread adoption of mobile phones as personal information artefacts worldwide[1], information centres and their staff are currently grappling with the impacts of this paradigm shift and how they effect the management of their spaces. As users participate in these spaces they are confronted with rules, policies or guidelines that are antithetical to their domesticated mobile phone practices. Being 'always on' and 'always with' their mobile phones symbolically and nominally imply that these devices have become an important component to how users seek, use, share and relate to information.

Yet information centres such as academic libraries have long traditions in offering and enforcing voice and noise-free spaces for information access [1]. Similar to notions of increasing integration of public and private spaces [2], academic libraries have become reinterpreted and rebranded as collaborative technology laboratories, information commons, and media labs used as both public gathering sites and as places for private and individual study [3]. In this study we explored the extent to which norms and policies within these hybridized information spaces acknowledge and/or make allowances for the use of oral communication both in face-to-face interactions and virtually via mobile phones. We investigated how historical understandings of users' (quiet) interactions with the information resources in academic libraries impact the way users engage with each other and with technology in these spaces for collaboration.

We conducted a documentary analysis of communication guidelines, policies, and posted signage regarding mobile phone use within an iSchool's integrated library and information studies laboratory (lab) to gain a contextual understanding of the manner in which user's communications practices were influenced by the language, messaging, and visual rhetoric embodied in these information sources. In addition, we conducted experiments over a two-week period in 2009 within the lab, involving both staged and observed mobile phone use within the space, and interviewed co-present staff and users to gather their perspectives on mobile phone use in the lab. Comparing the data from the documentary analysis with the experiment results we assessed influence of the official discourse on the structuring and facilitation of communicative interaction, and considered inherent contradictions in intent of the documented guidelines versus the expectation that the lab should be used as a collaborative space.

Early results indicated that the policy-makers within this information centre expressed hesitation in determining whether or not guidelines or policies should exist for mobile phone use in this hybridized information space, and what they should include. This led to a lack of explicit communication on mobile phone use leaving staff and users to interpret what was acceptable etiquette based on their experiences in other similar settings. Some users rapidly exited the space when they received an incoming call and expressed feelings of guilt about receiving a mobile phone call, particularly if the ring-tone was audible. Other users noted that since the space allows for face-to-face conversation among users that it is a space where mobile phone conversations may also take place; therefore, they felt that some use mobile phone use was to be expected. We observed staff using their mobile phones in the execution of their duties, and noted strategies that some users employed to strike what they believed to be an appropriate balance in reconciling feelings that mobile phone conversation was not allowed with their inclination to use the device as one of the information tools at their disposal.

Framing the analysis in science and technology studies we focus on the dynamic context within which these observations are made – a changing set of information practices involving mobile phones, a changing articulation of information centres as collaborative spaces, and the impact of embedded assumptions and expectations of how these spaces should be used based on historical precedents. We demonstrate that collaboration itself is interpreted in different ways when situated in specific types of environments and more so when face-to-face interaction is

[1] See for example research report by Kalba, Kas "The Global Adoption and Diffusion of Mobile Phones"[112 pages; December 2008], Information Resources Policy, Harvard University.

privileged over virtual collaboration via the mobile phone. We conclude with recommendations for policy-makers and managers of hybridized information spaces about how guidelines can be developed that involve a more in-depth understanding of mobile phone use in information spaces.

Keywords

mobile phones, cell phones, libraries, information centres, iSchool, space, place

REFERENCES

[1] Lever, K. and Katz, J. E. 2006. Cell phones in campus libraries: An analysis of policy responses to an invasive mobile technology. Processing and Management 43 (Mar. 2006), 1133-1139.

[2] Ling, R. 1997. On can talk about common manners: the use of mobile telephones in inappropriate situations. In Themes in Mobile Telephony, L. Haddon, Ed. Final Report of the COST 248 Home and Work Group.

[3] Barton, E. & Weismantel, A. 2007. Creating collaborative technology-rich workspaces in an academic library. Reference Services Review, 35, 395-404.

PhD Portal: Developing an Online iSchool Doctoral Student Community

Robin Naughton, Catherine Hall, Haozhen Zhao, Xia Lin
The iSchool at Drexel
College of Information Science and Technology
Drexel University
Philadelphia, PA 19104 USA
+1.215.895.2482
{chall, rnaughton, hzhao, xlin}@ischool.drexel.edu

ABSTRACT
iSchool doctoral students represent a diverse and growing community of information science and technology (IST) researchers. Each doctoral student manages his/her own IST career from admissions to graduation, using personal and varied methods of information management. As a result, lessons learned by students are not easily transmitted to each other. In this poster, we describe our design and development of an online doctoral student community that offers a single place for doctoral students to manage their daily graduate life from administration and course work to research and collaboration.

Categories and Subject Descriptors
H.5.2 **[INFORMATION INTERFACES AND PRESENTATION]**: User Interfaces - *User-centered design, Graphical user interfaces (GUI), Interaction styles*

General Terms
Design

Keywords
iSchool community portal, information management, user-centered design

1. INTRODUCTION

iSchool doctoral students juggle the demands of life and work on a daily basis by wearing many hats. They are students, research and teaching assistants, administrators, travel agents, presenters, friends, colleagues, collaborators, employees, parents, etc. At Drexel's iSchool, information regarding the doctoral student's IST career is distributed among many different technologies, making it difficult for doctoral students to manage their IST careers. According to Brooks and Fyffe (2004), "Undertaking doctoral studies is a huge shift in attitude and activity for most students" [1]. Thus, there is a strong need for a solution that offers a single place for doctoral students to manage their daily graduate life, whether it is course work, administration, research or socializing.

The iSchool PhD Experience is our design for a portal that connects doctoral students, whether near or far, helps them with research activities, guides them through administrative requirements, keeps them informed about what's happening in the field, and is a conduit for continuous communication and collaboration. It is built on a user-centered framework using Joomla [4] , an open-source content management system that offers flexibility, future growth and user development with little technical knowledge required for implementation.

2. BACKGROUND

In 2006, a group of IMLS Fellows in Drexel's iSchool, researched the need for a doctoral student community portal that could better support student research activities, administrative requirements and coursework. The team looked at available technology solutions, implemented an online survey to all IST doctoral students, and conducted a focus group with five students representing different stages of degree completion, and engaged in informal doctoral student group meetings and online discussions [3].

Their findings indicated that there was a strong need for a student community portal which could be one easy to use place for all information related to their doctoral studies. Results from the focus group indicated that the community portal needed to be an important part of the doctoral student program: "Of greater concern to the focus group was the motivation

required to have the system become a viable entity within the doctoral program" [3].

The research also identified constraints in terms of technologies. Based on these, the IMLS Fellows looked at open-source options that could support the needs of the doctoral student community.

This early research was the catalyst for a new group of IMLS Fellows to develop and implement the online doctoral student community in 2008.

3. DESIGN FRAMEWORK

Based on prior research, the online doctoral student community was designed with three important concepts in mind.

- **Resources** – students needed one place to find internal iSchool information (forms, FAQs, course offerings, etc.) as well as external information about the field (RSS Feeds, bookmarks, etc.).
- **Research** – students needed a way to communicate their research with each other, learn what is happening in their particular research area, and manage research projects online.
- **Community** – students needed a place to collaborate, communicate, and share information such as ideas, photos, videos, news, etc.

Some of the major features and functionality of the doctoral student portal include:

- Dynamic Content – user generated content can appear throughout the website in addition to automatically generated content such as RSS feeds of news, journals, blogs, etc.
- Faculty Research Tag Cloud – users can immediately learn about and identify iSchool Faculty research interests
- Forums – registered users can create discussion threads about courses, research, administration or just about anything.
- Projects – registered users can create and manage all aspects of a project and invite others to join the project through an online project management tool.
- Profile – registered users are automatically given a profile that can be edited and updated with personal and profession information. They can also create their own friendship networks by accepting and requesting friends within the community.

4. DEVELOPMENT

The development of the Drexel iSchool online doctoral student community focused on open-source technology solutions with flexibility and scalability. We analyzed the work of the previous IMLS Fellows and identified the important features and elements requested by the doctoral students and built a matrix that looked at each feature and its importance. Next, we reviewed the technology solutions suggested by the prior IMLS Fellows while looking at new open-source options available to determine the best solution for our needs. Technology options were eliminated based on compatibility with our servers, having the required features, time and skill required for implementation, flexibility, ease of use, and future scalability.

4. 1 Open-Source Options

The technology solutions were narrowed down to two open-source content management solutions with strong community support, Joomla [4] and Drupal [2], The matrix was updated to reflect these two options so that we could analyze each required feature with each possible solution.

Areas of Focus	Feature	Feature Elements	Drupal	Joomla	Notes	Priority
Community	Chat Client	Students can communicate with each other within the portal.	2	2		Low
Community	Group Activity	SIGs			Needs definition	High
Community	Group Activity	notification - if a new post occurred, users can get email if they opted in. Ability to opt in/out.	2	2		High
Community	Group Activity	RSS	1	1		High
Community	Group Activity	Project Management	3	3		High
Community	Calendar	Individual Calendar - ability to add events, links, etc.	5	3	Personal in some form on Joomla but not Drupal. Need to look at in detail for both systems.	Medium
Community	Calendar	General Calendar - ability to add events, links, etc.	2	2		High
Community	Calendar	Academic Calendar – school due dates, portfolio dates, conference deadlines, dissertation defense, form submission	2	2		High

Figure 1: Feature Matrix (Joomla & Drupal)

After analysis, we chose Joomla [4], which has a strong administrative focus that allows non-developers to quickly install and manage the system. It also offered a large number of modules that can be easily installed to meet the needs of the community.

4. 2 Online Doctoral Student Community

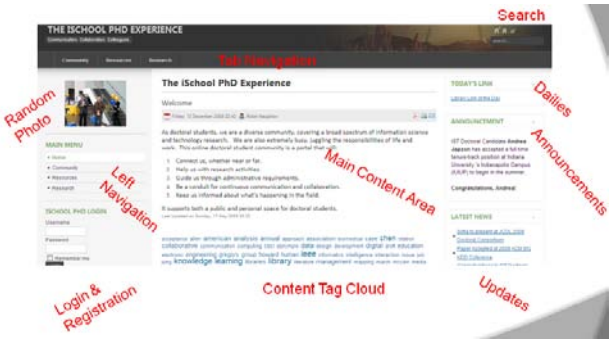


Figure 2: The iSchool PhD Experience home page

The online iSchool doctoral student community is now in its early stages. New and current doctoral students are joining the

community and contributing to grow and develop the community to meet their personal and group needs.

4. 3 Content

Based on the analysis of the user's needs, we developed content in two major areas. One is the resources for the process of doctoral studies. These include the academic calendars, the PhD program descriptions and requirements, various forms for portfolio reviews, advance to candidacy, and graduation, etc. The portal provides one place to access all these materials. The other resource is for the research need of doctoral students, which can help to develop a learning community [5]. These include access to all the publications of the ischool faculty, major readings of key authors in the field, research activities in other iSchools, and recent articles in relevant journals (such as JASIST and ARIST). Selected Bookmarks on relevant topics from Delicious and CiteULike are also included.

Most of the content are either from the user's input or the RSS feeds and other dynamic links. We expect the content to grow vigorously as the user community grows.

5. CONCLUSION & FUTURE WORK

The PhD Portal helps doctoral students with their IST careers and provides a much needed place for them to collaborate and communicate online.

The work on the portal will continue. The future work includes analyzing current system use and functionality to determine benefits and possible feature upgrades, expanding the portal to doctoral students in other iSchools beyond Drexel, adding requested features, and developing a method for continued growth and development of the community.

6. REFERENCES

[1] Brooks, C. & Fyffe, J. (2004). Are we comfortable yet? Developing a community of practice with PhD students at the University of Melbourne. In R. Atkinson, C. McBeath, D. Jonas-Dwyer & R. Phillips (Eds), *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (pp. 163-169). Perth, 5-8 December. http://www.ascilite.org.au/conferences/perth04/procs/brooks.html

[2] Drupal (n.d.). Drupal home page. http://drupal.org/

[3] IMLS Fellows. (2006). *Developing an Online Portal to Support Doctoral Student Activities.* (Unpublished).

[4] Joomla (n.d.). Joomla home page. http://www.joomla.org/

[5] Twale, D. , Korn, K. A., Shafer, C. and Hibner, K. 2008-10-15 "Building Online Learning Community for Doctoral Students" *Paper presented at the annual meeting of the MWERA Annual Meeting, Westin Great Southern Hotel, Columbus, Ohio Online* <**PDF**>. 2009-05-23 from http://www.allacademic.com/meta/p273523_index.html

Commonality Analysis: Demonstration of an SPSS Solution for Regression Analysis

Kim Nimon, Ph. D.
University of North Texas
College of Information
3940 N. Elm, Suite G150
Denton, TX 76207
kim.nimon@unt.edu

Mariya Gavrilova
University of North Texas
College of Information
3940 N. Elm, Suite G150
Denton, TX 76207
mariya.gavrilova@unt.edu

ABSTRACT
Multiple regression is a widely used technique to study complex interrelationships among people, information, and technology. In the face of multicollinearity, researchers encounter challenges when interpreting multiple linear regression results. Although standardized function and structure coefficients provide insight into the latent variable ($\hat{Y}$) produced, they fall short when researchers want to fully report regression effects. Regression commonality analysis provides a level of interpretation of regression effects that cannot be revealed by only examining function and structure coefficients. Importantly, commonality analysis provides a full accounting of regression effects which identifies the loci and effects of suppression and multicollinearity. Conducting regression commonality analysis without the aid of software is laborious and may be untenable, depending on the number of predictor variables. A software solution in SPSS is presented for the multiple regression case and demonstrated for use in evaluating predictor importance.

Categories and Subject Descriptors
G.3 [Probability and Statistics]: Correlation and regression analysis

General Terms
Theory, Measurement

Keywords
Commonality analysis, multicollinearity, suppression

1. COMMONALITY ANALYSIS
Developed in the 1960s as a method of partitioning variance (R^2) [4],[5],[6],[7], commonality analysis provides a method to determine the variance accounted for by respective predictor variables [9],[11]. Commonality analysis partitions a regression effect into unique and common effects. Unique effects identify how much variance is unique to an observed variable, and common effects identify how much variance is common to groups of variables. The number of equations required for a commonality analysis is 2^k-1 components, where k is the number of predictor variables in the regression analysis. The sum of unique and common effects equals the total variance in the dependent variable explained by the predictor variables. For a detailed discussion of commonality analysis, readers are encouraged to consult [8].

2. ILLUSTRATIVE EXAMPLE
To illustrate the benefit of commonalty analysis, an example is provided from Yao, Rice, and Wallis [12]. Yao, Rice, and Wallis examined how need for privacy (PrivNeed), self-efficacy (SelfEff), beliefs in privacy rights (PrivRight), Internet use diversity (DivUse), Internet use fluency (FlueUse) related to online privacy concerns (PrivConc). They used a two step hierarchical regression analysis where Internet use diversity and Internet use fluency were first regressed on the dependent variable. In the second step, they entered the three psychological and belief variables. Although they indicated they wanted to examine the unique effects of each of the hypothesized factors, their analyses actually indicated how much variance the three psychological and belief variables contributed to variability in concerns about online privacy after controlling for the effects of Internet use diversity and Internet use fluency. We conducted commonality analysis based on the correlation data from their study to demonstrate its analytic capability and to answer the researchers' identified research question.

3. SOFTWARE DEMONSTRATION
To perform the regression commonality analysis, we used an SPSS script that was developed based on the R code published by Nimon, Lewis, Kanes, and Hayes [8] that conducts commonality analysis for any number of predictor variables. The SPSS script file can be obtained at not cost by contacting the lead author. As depicted in Figures 1 – 4, the script file prompts the user for the: (1) SPSS data file, (2) output filename prefix, (3) dependent variable, and (4) independent variables. Due to limitations in the SPSS MATRIX command, all variables names must be eight characters or fewer.

Using the information supplied, the script generates two SPSS data files – CommonalityMatrix.sav and CCByVariable.sav where both file names are prepended with the output file name prefix. CommonalityMatrix.sav contains the unique and common commonality coefficients as well as the percent of variance in the regression effect that each coefficient contributes. The individual entries in the table can be used to determine how much variance is explained by each effect as well as which coefficients contribute most to the regression effect. CCByVariable.sav provides another view of the commonality effects. The unique effect for each of the predictors is tabularized, as well as the total of all common effects for which the predictor is involved. The last column sums the unique and common effects. Dividing the variance sum by the regression effect yields the percent of variance explained by each variable, equivalent to a squared structure coefficient. The benefit

of employing commonality analysis in conjunction with the analysis of squared structure coefficients is that the researcher can determine how much variance each variable uniquely contributes and how much each shares, if any, with every other variable in the regression [8].

Based on the example, Tables 1 and 2 respectively contain the contents of YaoCommonalityMatrix.sav and YaoCCByVariable.sav. Table 3 presents an example of how the commonality effects by variable can be displayed alongside traditional multiple regression output to add another layer of consideration when evaluating the importance of predictors.

4. COMMONALITY INTERPRETATION

In Yao, Rice, and Wallis [12], the majority of the regression effect was explained by variance that was unique to belief in privacy rights (61.63%), need for privacy (14.61%), and self-efficacy (3.06%). Internet use diversity and fluency contributed little unique variance to explaining differences in online privacy concerns. In total, the four predictors uniquely accounted for 80.290% of the regression effect. The remaining 19.710 was due to variance the sets of predictors shared in common with the dependent variable. The most noticeable common effect was between need for privacy and beliefs in privacy rights, which accounted for 9.55% of the regression effect.

The commonality coefficients further indicate the presence of negative commonality coefficients. Negative commonalities occur in the presence of suppressor effects when some of the independent variables affect each other in the opposite direction [10]. While Frederick [3] indicated that negative commonalities should be interpreted as zero, others have disagreed [1], [2], [10]. Negative commonality coefficients indicate the amount of variance in the regression effect that is confounded by a set of predictor variables. In the case of suppression, negative commonality coefficients identify the increase in power associated with the suppressor effect. The commonalty data in Table 1 indicate that the regression effect was confounded by 7 out of the 15 predictor variable combinations involving self-efficacy. Suppression accounted for 3.049% of the regression effect.

The data in Table 3 demonstrate the benefits of fully reporting regression effects. In one table, researchers can simultaneously consider beta weights, structure coefficients, unique effects, and common effects when evaluating the importance of predictors. For example, while Internet use fluency might be considered an unimportant predictor due to its insignificant beta weight, its squared structure coefficient indicates that it explains a moderate amount of the regression effect. The discrepancy between the significance of the variable's beta weight and its contribution to the regression effect can easily be explained as most of its effect is due to variance that it shares in common with other predictor(s). On the other hand, the data in Table 3 demonstrates that the agreement between the relative importance of beliefs in privacy rights based on its beta weight and structure coefficient is due to the magnitude of unique variance that the variable contributes to the regression effect.

5. CONCLUSION

From a didactic perspective, commonality analysis clarifies the roles that multicollinearity and suppression play in the relationship between standardized function and squared structure coefficients. In addition, it can be observed that commonality analysis subsumes the role of computing squared structure coefficients because the portion of the regression effect explained by each variable generated from the canonical commonality analysis is identical to the squared structure coefficient generated from multiple linear regression. From a theoretical perspective, regression commonality analysis can provide important insights into variable relationships.

6. REFERENCES

[1] Beaton, A. E. 1973, March. Commonality. (ERIC Document Reproduction Service No. ED111829)

[2] Capraro, R. M., and Capraro, M. M. 2001. Commonality analysis: Understanding variance contributions to overall canonical correlation effects of attitude toward mathematics on geometry achievement. Multiple Linear Regression Viewpoints, 27(2), 16-23.

[3] Frederick, B. N. 1999. Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), Advances in social science methodology, 5, 305-318. StaMford, CT: JAI Press.

[4] Mayeske, G. W., Cohen, W. M., Wisler, C. E., Okada, T., Beaton, A. E., Proshek, J. M., Weinfeld, F. D., and Tabler, K. A. 1969. *A study of our nation's schools.* Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

[5] Mood, A. M. 1969. Macro-analysis of the American educational system. Operations Research, 17, 770-784.

[6] Mood, A. M. 1971. Partitioning variance in multiple regression analyses as a tool for developing learning models. American Educational Research Journal, 8, 191-202.

[7] Newton, R. G., and Spurrell, D. J. 1967. A development of multiple regression for the analysis of routine data. Applied Statistics, 16, 51-64.

[8] Nimon, K., Lewis, M., Kane, R., and Haynes, R. M. 2008. An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. Behavior Research Methods, 40, 457-466. DOI= 10.3758/BRM.40.2.457

[9] Onwuegbuzie, A. J., and Daniel, L. G. 2003, February 19. Typology of analytical and interpretational errors in quantitative and qualitative educational research. Current Issues in Educatio*n* [On-line], 6(2). Available: http://cie.ed.asu.edu/volume6/number2/

[10] Pedhazur, E. J. 1997. Multiple regression in behavioral research: Explanation and prediction (3rd ed.). Ft. Worth, TX: Harcourt Brace.

[11] Rowell, R. K. 1996. Partitioning predicted variance into constituent parts: How to conduct regression commonality analysis. Advances in Social Science Methodology, 4, 33-43.

[12] Yao, M. Z., Rice, R. E., and Wallis, K. 2007. Predicting user concerns about online privacy. Journal of the American Society for Information Science and Technology, 58, 710-722. DOI= 10.1002/asi.20530

Table 1. YaoCommonalityMatrix

Variables	Coefficient	%Total
Unique to PrivNeed	0.030	14.605
Unique to SelfEff	0.006	3.059
Unique to PrivRght	0.127	61.529
Unique to DivUse	0.001	0.444
Unique to FlueUse	0.001	0.654
Common to PrivNeed SelfEff	0.008	3.804
Common to PrivNeed PrivRght	0.020	9.555
Common to SelfEff PrivRght	-0.001	-0.390
Common to PrivNeed DivUse	0.000	0.116
Common to SelfEff DivUse	-0.001	-0.247
Common to PrivRght DivUse	0.002	0.914
Common to PrivNeed FlueUse	0.001	0.343
Common to SelfEff FlueUse	-0.001	-0.263
Common to PrivRght FlueUse	0.006	2.756
Common to DivUse FlueUse	0.001	0.505
Common to PrivNeed SelfEff PrivRght	0.002	1.029
Common to PrivNeed SelfEff DivUse	0.000	-0.139
Common to PrivNeed PrivRght DivUse	0.001	0.254
Common to SelfEff PrivRght DivUse	0.000	-0.189
Common to PrivNeed SelfEff FlueUse	0.000	-0.182
Common to PrivNeed PrivRght FlueUse	0.002	1.024
Common to SelfEff PrivRght FlueUse	-0.001	-0.334
Common to PrivNeed DivUse FlueUse	0.000	0.204
Common to SelfEff DivUse FlueUse	-0.001	-0.219
Common to PrivRght DivUse FlueUse	0.004	1.672
Common to PrivNeed SelfEff PrivRght DivUse	0.000	-0.203
Common to PrivNeed SelfEff PrivRght FlueUse	-0.001	-0.287
Common to PrivNeed SelfEff DivUse FlueUse	0.000	-0.162
Common to PrivNeed PrivRght DivUse FlueUse	0.001	0.583
Common to SelfEff PrivRght DivUse FlueUse	0.000	-0.155
Common to PrivNeed SelfEff PrivRght DivUse FlueUse	-0.001	-0.279
Total	0.207	100.000

Table 2. YaoCCbyVariable

Variable	Unique	Common	Total
PrivNeed	0.030	0.032	0.063
SelfEff	0.006	0.004	0.010
PrivRght	0.127	0.033	0.160
DivUse	0.001	0.006	0.006
FlueUse	0.001	0.011	0.012

Table 3. Regression Results for Yao, Rice, and Wallis (2007) Data Predicting Online Privacy Concerns

Predictor	*R*	R^2	R^2_{adj}	*β*	*p*	*Unique*	*Common*	*Total*	*% of R2*
	0.454	0.206	0.197						
PrivNeed				0.179	<0.001	0.030	0.032	0.063	0.303
SelfEff				-0.083	0.073	0.006	0.004	0.010	0.049
PrivRght				0.366	<.001	0.127	0.033	0.160	0.777
DivUse				0.033	0.493	0.001	0.006	0.006	0.031
FlueUse				0.040	0.406	0.001	0.011	0.012	0.059

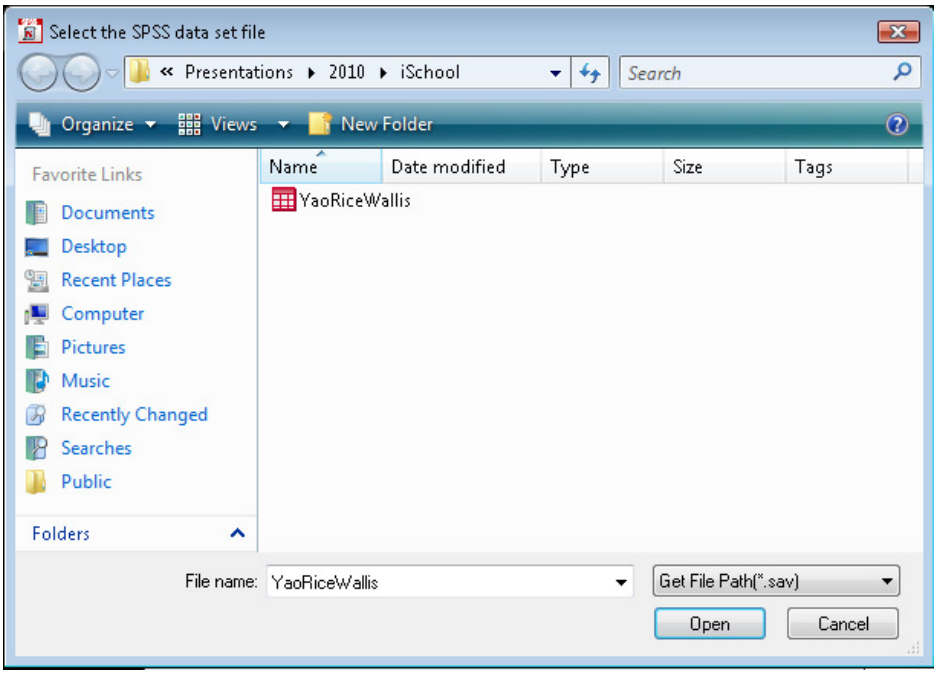


Figure 1. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 1.

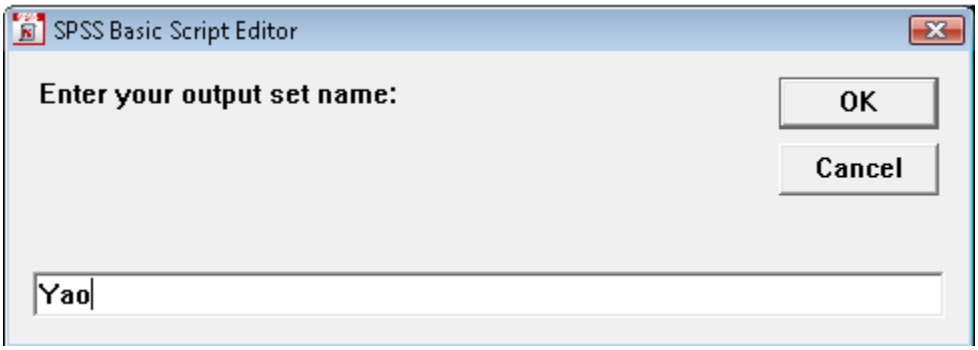


Figure 2. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 2.

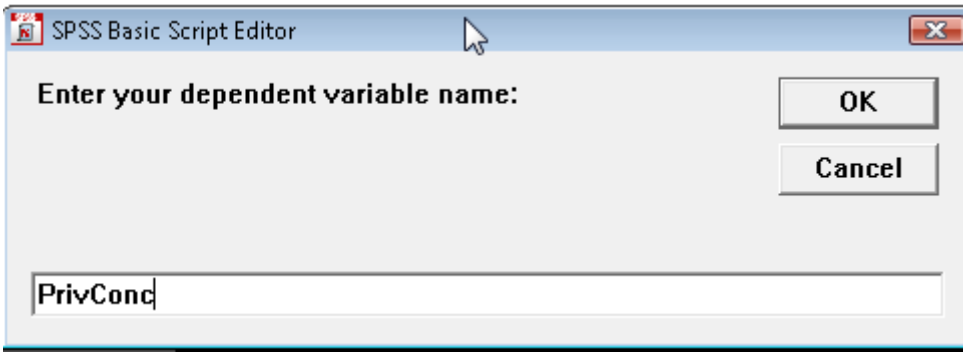


Figure 3. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 3.

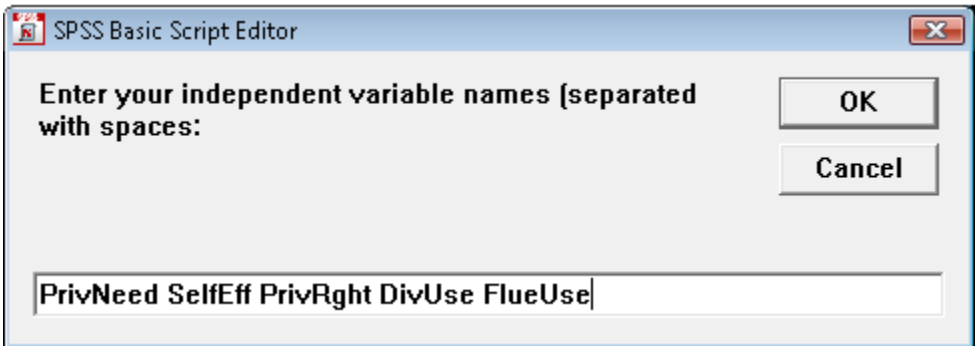


Figure 4. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 4.

Outside the Frame: Modeling Discontinuities in Video Stimulus Streams

Richard L. Anderson
Visual Thinking Laboratory
College of Information
University of North Texas
Denton, Texas
1.940.206.1172
rich.anderson@unt.edu

Brian C. O'Connor
Visual Thinking Laboratory
College of Information
University of North Texas
Denton, Texas
1.940.206.1172
brian.oconnor@unt.edu

Melody J. McCotter
College of Information
University of North Texas
Denton, Texas
1.940.206.1172
melody.mccotter@unt.edu

ABSTRACT

How are we to get beyond the literary metaphor Augst asserts is central problem with film analysis? How are we to step outside the "shot" as the unit of analysis - the "shot" which Bonitzer claims is useless for analysis because of researchers' "endlessly bifurcated" definitions of "shot?

We have had success with a form of computational structural analysis which incorporates the viewer into the model. Comparing changes in levels of Red, Green, and Blue from frame to frame and comparing the patterns of change with an expert film theorist's model.

We are currently analyzing discontinuities in the entire data stream of a film. We are asking just what aspects of the data stream account for viewer reactions. We are examining distribution of color, edges, luminance, and other components. By modeling changes in the various stimuli over time within a vector space model and comparing those changes with the responses of (at first) an expert viewer, then with a variety of viewers we should be able to make strides in matching forms of representation to the most effective mode of representation for the individual user; and at the same time provide a set of analytic tools that account for the multiple time-varying signals that make up a movie, whether a cell phone video or Hollywood blockbuster.

Significantly, we now step outside the frame as the unit of analysis and look to the possibilities of analysis at the sub pixel level. That is, analysis of one component of a pixel location such as luminance or merely the green component (no red or blue provides a very fine grained level of examination. At the same time, the vector space model provides a way of examining the stimulus effect of multiple threads that do not necessarily change in synch.

As we consider these possibilities, we begin to see a general model of a document as a continuous stream of data that either (as a whole or in part) functions as a stimulus or does not.

Our poster will present graphical representations of changes in the data stream for the "Bodega Bay" sequence of Hichcock's THE BIRDS and the reactions of Raymond Bellour, whose analyses and modeling of Hichcok's works and of classic Hollywood film in general are held in high regard. We begin with Bellour and the Bodega Bay sequence because we have already published research on this data and, thus, have a significant foundation upon which to build. We will then apply the same techniques to a set of other works.

Categories and Subject Descriptors

H.3.1 [**Content Analysis and Indexing**]: Abstracting Methods

H.3.3 [**Information Search and Retrieval**]: Retrieval Models

I.2.4 **[Knowledge Representation and Methods]**

Keywords

Video, Key Frames, Document Theory, Information Theory, Functional Ontology

The Use and Misuse of Science: Refining the Theoretical Framework of Science Policy

Shannon M. Oltmann
School of Library & Information Science
Indiana University
1310 E 10th Street, LI 001
Bloomington, IN 47404
soltmann@indiana.edu

1. ABSTRACT

This poster examines the use and misuse of science information in the federal government. Scientific information is a vital component of policy making in the U.S. today. Stine notes that science research is "intricately linked to societal needs and the nation's economy in areas such as transportation, communication, agriculture, education, environment, health, defense, and jobs" [7, p. i]. In the past, the relationship between science and policy was seen as a linear process: science conducted research, collected data, and presented its findings to federal agencies, which then use that evidence to determine the best policy action [2, 5].

However, the reality of science policy is far more complex; while science is a valuable source of information, it is also problematic, since scientific data may conflict with political, moral, and economic values [5, 6, 7]. For example, if endangered fish reside in a lake, politicians may face choices between preserving the ecosystem, irrigating nearby farms, and allowing recreational use of the lake. Each choice has economic, environmental, and political ramifications. Doremus explains that "esthetic, ecological, educational, historical, recreational, or scientific" values can all be considered relevant foundations for agency decisions [3, p. 1136]. Because of this complexity, "the political community and the scientific community… collaborate at the boundary of politics and science over the integrity and productivity of research" [5, p. 143]. In this conceptualization, "government cannot make good policy decisions unless the decision makers have access to, and appropriately use, the best available understanding of the facts" [4, p. 1639].

Federal agencies, like individuals, have information behaviors—they create, access, review, share, evaluate, and act upon information in order to formulate and assess public policy.

Agencies could accept scientific conclusions and use them as the basis of policy formation. Agencies could accept the science, yet determine that it is not the best or sole basis of effective policy. Of course, agencies could reject or partially reject the science, thus creating more opportunities to basis policy on other considerations. Typical agency behavior with respect to science falls across a spectrum, with science being neither unreservedly endorsed nor discarded. While "a scientist views science as a way of learning, a policy maker…may see science as the justification for a decision, a requirement of the law, a tool or impediment, or something that opposes or supports their viewpoint" [1, p. 1005]. Furthermore, agency information behavior with respect to science does not exist in isolation. There is recurring interaction between science and policy. For instance, scientists who study the toxic effects of chemicals and report their conclusions to the Environmental Protection Agency, to guide agency behavior, will likely continue studying the same chemicals and providing additional information to further influence policy. How the EPA behaves with respect to the scientific information may shape future research, communication efforts, or the information behaviors of the scientists themselves.

Principal-agent theory is frequently used to explain how science and policy interact. Under this approach, federal agencies, as principals, contract with science to provide needed information. Science then acts as an agent, supplying data and conclusions in exchange for funding, prestige, and other rewards [5, 6]. Principal-agent theory captures a significant portion of the interaction between science and policy, but does not reflect the entire relationship. Specifically, principal-agent theory has little to say about how agencies use science—the information behaviors in which they engage—or how these information behaviors affect subsequent interaction with scientists. The theory currently does not address the problem of under-utilized or under-appreciated agents. If the agents perceive their work is not incorporated into policy, perhaps they will refuse to do further work, will begin doing shoddy work, or will attempt to subtly integrate policy advice into their work. Since these information behaviors are, in fact, a crucial part of formulating policy, they ought not be overlooked.

The nature of the recurring interactions, and how they are affected by agencies' information behavior, has not been explicitly examined in the previous literature. This poster illuminates these aspects of the relationships between science and policy. Specific

examples of agencies using and misusing scientific information will be drawn from the literature to illustrate the complex interactions. The full, cyclical relationship between science and policy will be portrayed, demonstrating how agencies' information behaviors may affect subsequent research and communication behaviors. This will necessarily entail a refinement of principal-agent theory as it has been applied to science policy.

This research will be a valuable contribution in several ways. It brings science policy—how scientific information is used or misused—to the explicit attention of iSchools and their cognate fields of study. As we create technological tools and engage in policy-relevant research, we need to pay attention to how our data and conclusions may or may not be utilized. In addition, science policy can benefit from the theoretical and conceptual rigor of the trans-disciplinary research of the iSchools. Finally, the research will also test and strengthen the use of principal-agent theory as it applies to science policy. Overall, this theory has great utility, but can be refined to address more of the interaction between science and policy.

2. References

[1] Brosnan, D.M. (2007). Science, law, and the environment: The making of a modern discipline. *Environmental Law, 37*, 987-1006.

[2] Bush, V. (1945). *Science: The endless frontier.* Office of Scientific Research & Development. Retrieved June 1, 2008 from http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm

[3] Doremus, H. (1997). Listing decisions under the Endangered Species Act: Why better science isn't always better policy. *Washington University Law Quarterly, 75*, 1029-1153.

[4] Doremus, H. (2008). Scientific and political integrity in environmental policy. *Texas Law Review, 86*, 1601-1653.

[5] Guston, D.H. (1999). *Between politics and science: Assuring the integrity and productivity of research.* Cambridge: Cambridge University Press.

[6] Guston, D.H. (2003). Principal-agent theory and the structure of science policy, revisited: 'Science in policy' and the US Report on Carcinogens. *Science & Public Policy, 30*(5), 347-357.

[7] Stine, D. (2008, April 22). *Science and technology policymaking: A primer.* CRS Report for Congress: RL 34454.

Libraries as Bridges across the Digital Divide: Partnerships and Approaches Used in the U.S. Technology Opportunities Program, 1994-2005

Anna Pederson
GSLIS/University of Illinois
501 E Daniel St, MC-493
Champaign, IL 61820
217-244-0263
apeders5@illinois.edu

Kate Williams
GSLIS/University of Illinois
501 E Daniel St, MC-493
Champaign, IL 61820
217-244-9128
katewill@illinois.edu

ABSTRACT

The purpose of the poster is to show how libraries used government funds and community partnerships to close the digital divide in the United States. Part of the mission of libraries is to bridge the digital divide. As an answer to the digital divide, the U.S. government started a grant program in 1994. Over ten years, the Technologies Opportunities Program (TOP) awarded $230 million to 600 communities to promote network technology and community partnership.

The digital divide is a rich concept rather than a simple binary divide. It's something that is nuanced, multidimensional and ever-changing. Everyone is immersed in the digital divide in one respect or another because none of us are on the same plane of learning and expertise. We have learned much from the plethora of research that has taken place in communities in the United States and abroad. This study sheds like on the digital divide and how libraries have addressed it.

Of the 600 projects funded by TOP, 25 were library-led: approximately 10 took place in public libraries, three in academic libraries, and 12 in library networks or other settings. This research uses the TOP Data Archive, which we created with the help of others including the U.S. Department of Commerce itself, to examine these 25 projects. We have constructed tables and word clouds to find trends and analyze the projects and partnerships and will use established network analytical methods as well. Interviews with key leaders in each of the projects will help ascertain how each project developed over time. Our governing theory is that social capital and social networks contribute to ICT use.

Our questions include: How did the partnerships between the library and other organizations affect each project? How did they define success, and did they achieve it? Our first finding is that libraries adapted the grant program to their own strategic activities and did not set library work aside. Second, the libraries took three main approaches: to build computer networks with wires and fiber-optics, to build the human-computer infrastructure known as a Freenet, or to create new library programs to help their community use technology. We will also present data on the programs and the size and shape of the partnerships that carried them out.

Our research has found a total of 80 partnerships across 25 separate library-led TOP projects. Each project had an average of 4.3 partnerships; with the maximum being 11 and the minimum number of partnership being one. Our analysis included a typology of partners: education, corporations, government, and organization. Educational partners include schools, colleges, universities, and other educational organizations. Corporate entities are defined as businesses or companies. Government partners maybe municipal, city, state or national government entities. Lastly, organization is a broad category that fits every type of non-profit organization, whether it be community, environmental, educational, etc. There are also four sub-categories: library, health, art, and communications. Library partners may be local, state, college, or university libraries. Health institutions are any health organization, whether government or community, or hospitals. Art partners involve art museums, local art organizations, etc. Lastly, communications partners are communication corporations, TV or radio stations, or government communication entities. The categories will allow us to investigate the relationship between the type of partners in each project and the scope and outcome of each project. The data includes 33 education partners, 28 government, 23 organization, 8 libraries, 8 communications 6 corporation, 5 health, and 3 art.

In December and January we will use NetDraw to create a visual representation of the egocentric network of a library and its partners, and look for patterns. We will also carry out telephone interviews with the leaders of each project. The phone interviews will tell us about long-term projects outcomes and how the partnerships advanced or impeded each project.

This poster will provide insights and suggestions to libraries that are working on the digital divide or on building partnerships. Since the U.S. has yet to catch up with the rest of the world in terms of broadband speed and utilization, the government has started another round of grants called the Broadband Technology Opportunities Program; our findings will also inform that work. Libraries have the responsibility to serve increasingly disparate populations and our poster provides an analysis of an important group of library projects which have never been presented to an international audience. This topic will be of interest to many people in the library profession, especially those dedicated to serving the public through the use of innovative technology.

Relevant links:

TOP archive at the University of Michigan: http://quod.lib.umich.edu/cgi/f/findaid/findaid-idx?c=sclead&idno=umich-spc-Power-Top

Broadband Technology Opportunities Program: http://www.ntia.doc.gov/broadbandgrants/

Kate Williams' lecture on the digital divide: http://www.lis.illinois.edu/news/digital_divide09.html?videoID=bj9KpV0jak2KjRozoRflhA

Folktales & Folksonomies: Investigating the Utility of Tags as a Means of Description for Folktales

Pirmann, Carrie M.
University of Illinois at Urbana-Champaign
pirmann2@illinois.edu

Topics: Information organization, Social tagging, Folksonomies
Keywords: folksonomy, tagging, information access, folktales

This research considers the utility of folksonomies as a means of augmenting access to a collection of folktale resources. Catalog records for item classed as folktales often contain sparse information about facets of the stories (e.g., characters, setting, cultural or geographic origins of a tale, moral, motif) that may be valuable to users of these resources. The Folktales, Facets, & FRBR Project[1] seeks to improve access to items in the Center for Children's Books (CCB) folktales collection through the enhancement of existing catalog records for these items. Folksonomies may serve as one means of augmenting records for this collection. Facet analysis may aid in identifying areas in which folksonomies may provide particularly rich sources of data for enhancing item records.

Prior work with the CCB folktales books included a facet analysis of 100 titles[2] from the collection. From this analysis, an initial set of 12 top-level categories was established. Catalog records for the 100 items were then analyzed with these facet categories serving as a framework for the analysis. It was found that certain facets (e.g., author, type of story, cultural or geographic origins of a story) were well represented in the standard bibliographic data found in catalog records. Conversely, information about characters, setting, and audience was very sparsely represented in the records. Subject information was found in approximately one third of the records. Facets such as motif, moral, and illustrations were not at all represented.

To examine the use of folksonomies with the CCB folktales collection, LibraryThing for Libraries was installed into the collection's Koha (open source ILS) OPAC. LibraryThing for Libraries[3] is a series of bibliographic enhancements that display LibraryThing tags on existing catalog records. Catalog records from 100 books

---

[1] Currently underway at the Center for Children's Books, Graduate School of Library & Information Science, University of Illinois at Urbana-Champaign; Professors Kathryn La Barre and Carol Tilley are the principal co-investigators. More information is available at: http://cirss.lis.illinois.edu/CollMeta/Folktales.html

[2] There are approximately 1500 items total in the CCB folktales collection.

[3] More information on LibraryThing for Libraries can be found at: http://www.librarything.com/forlibraries/

from in the collection were examined, and all LibraryThing tags displayed on each record were recorded in a database[4]. Facet analysis was conducted on these tags, with the facet categories established in the earlier analysis providing a guide for the present analysis.

Initial results of the analysis indicated that the folksonomic tags most frequently fell into the facet categories of type of story, subject, characters, geographic and cultural references, and type of book. In particular, tags provided better descriptive data for the facets of subject, characters, and type of story. It was observed that the ambiguity of tags may be problematic for categories such as geographic and cultural references (e.g., the tag "Russia" may indicate the country of origin of a story, or its setting). The analysis of the LibraryThing tags also revealed some additional categories (e.g., mood, textual elements) not found in the initial facet analysis performed on the 100-book sample. Further analysis of LibraryThing tags on items in the CCB folktales collection may be of use to draw out additional facet categories that would enhance access to items in this collection.

[4] Due to the limitations of LibraryThing for Libraries, only 59 titles had any tag data displayed on their records.

The Influence of Document Indexing on the Bilinear Property of Vector Space Model

Peng Qu
Department of Information Management, Peking University
Beijing 100871, China
pqu@pku.edu.cn

Abstract

The paper discusses on the influence of TFIDF indexing method on inner product and bilinear function which are fundamental for Vector Space Model; and comes to the conclusion that indexing process can change the bilinear property of inner product space. It also comes to the finding that the maximum term frequency influences vector space greatly; and that the normalization factor has both retrieval and geometrical values.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models

General Terms

Algorithms, Theory.

Keywords

Indexing, Vector Space Model, Inner Product

1. INTRODUCTION

Cater discusses the norm p's influences on information retrieval space. He proves that information retrieval systems with different weights and norm values are homeomorphous in topology [1]. Researchers using topological methods have made many valuable attempts since then. Representatives of such researches come from Egghe and Rousseau [5, 6] and Dominich's research on Information Retrieval Framework. Dominich proves that the similarity information retrieval on E is equivalent to the Hausdorff space inducted by $\delta = 1 - \sigma$ on E [2, 3]. This shows that information retrieval space is good in topology and provides the foundation for further development of information retrieval theory.

Van Rijsbergen uses a totally different method and explicates that algebraic information retrieval spaces can be regarded as a Hilbert space [10]. These good characters of vector space show that information retrieval space is theoretically proper and self-contained. However, if we examine carefully his deductions and argumentations, we find his assertion is based on the assumption that document vector can substitute document totally. Actually, this is also the problem with many information retrieval models with algebraic framework. Dominich and Kiezer point out that vector space model is only related to data structures, and that the definition of inner product is operation on data structures rather than logical or formal tools [4].

Like Dominich and Kiezer [4], we will not take the assumption that document vectors can simply represent documents. Document vectors are the results of indexing and weighting. To neglect the difference is to neglect indexing process. Starting from the difference, we define the indexing function α to explicate indexing process formerly hidden in information retrieval models. We differentiate document d from document vector $\vec{d}$; and bridge the two with indexing function α. We will examine the influence of indexing on VSM's inner space in the paper.

2. INDEXING PROCESS INFLUENCE ON DOCUMENT REPRESENTATION

Given document set **D**, indexing term set **T**, documents $d_1, d_2, \cdots, d_m$, in which, $m = |\mathbf{D}|$. The indexing function α is defined as

$$\alpha(d_i) = \overrightarrow{d_i} = (w_{i1}, w_{i2}, \cdots, w_{in})'$$

in which, $n = |\mathbf{T}|$, w_{ij} is the weight of indexing term $t_j \in \mathbf{T}$ in document d_i. According to standard TFIDF weighting formula, we have

$$w_{ij} = tf_{ij} \cdot idf_j = \frac{Freq_{ij}}{\max_i Freq_{ij}} \log \frac{n}{n_j}$$

$Freq_{ij}$ is the raw frequency of indexing term t_j in document d_i. n_j is the number of documents in which indexing term t_j appears. Define the mapping from document set **D** and query set **Q** to real number field as match function

$$\rho: \mathbf{D} \times \mathbf{Q} \to \mathbf{R}$$

It is thus accordant with the similar match in Vector Space Model. Meanwhile, for the present information retrieval systems, we can regard the document set **D** and query set **Q** are approximately identical. It thus satisfies the basic requirement of inner product formally. For $d_1 \in \mathbf{D}$, $q \in \mathbf{Q}$, we have

$$\rho(d_1, q) = \alpha(d_1)\vec{q} = \overrightarrow{d_1}'\vec{q} = (w_{11}, w_{12}, \cdots w_{1n})\vec{q}$$

We examine the linearity on document $d_i \in \mathbf{D}$ now. For $\forall d_1, d_2 \in \mathbf{D}$, we have

$$\rho(d_1 + d_2, q) = (w_{(1+2)1}, w_{(1+2)2}, \cdots w_{(1+2)n})\vec{q}$$

Noticing that we cannot simply add the document vectors $\overrightarrow{d_1}, \overrightarrow{d_2}$ together, we take $d_1 + d_2$ as one document. For any member $w_{(1+2)j}$ in $\overrightarrow{d_1 + d_2} = \alpha(d_1 + d_2)$, we have

$$w_{(1+2)j} = \frac{Freq_{(1+2)j}}{\max\limits_{1+2} Freq_{(1+2)j}} \log \frac{n-1}{n'_j}$$

in which, $n'_j = n_j$ or $n'_j = n_j - 1$, relating to whether indexing term t_j occurs in both d_1, d_2 or not. When the database is large enough, we can regard

$$\log \frac{n-1}{n'_j} \cong \log \frac{n}{n_j}$$

Noticing that $Freq_{(1+2)j}$ is the sum of raw frequencies of documents d_1 and d_2, the above formula can be expanded as

$$w_{(1+2)j} = \frac{Freq_{1j} + Freq_{2j}}{\max\limits_{1+2} Freq_{(1+2)j}} \log \frac{n}{n_j}$$

i.e.

$$w_{(1+2)j} = \frac{\max\limits_{1} Freq_{1j}}{\max\limits_{1+2} Freq_{(1+2)j}} \frac{Freq_{1j}}{\max\limits_{1} Freq_{1j}} \log \frac{n}{n_j} + \frac{\max\limits_{2} Freq_{2j}}{\max\limits_{1+2} Freq_{(1+2)j}} \frac{Freq_{2j}}{\max\limits_{2} Freq_{2j}} \log \frac{n}{n_j}$$

We thus have

$$w_{(1+2)j} = k_1 w_{1j} + k_2 w_{2j}$$

in which

$$0 < k_i = \frac{\max\limits_{i} Freq_{ij}}{\max\limits_{1+2} Freq_{(1+2)j}} \leq 1, i = 1,2$$

Thus

$$\begin{aligned} &\rho(d_1 + d_2, q) \\ &= (k_1 w_{11} + k_2 w_{21}, k_1 w_{12} + k_2 w_{22}, \cdots, k_1 w_{1n} + k_2 w_{2n})\vec{q} \\ &= k_1 (w_{11}, w_{12}, \cdots, w_{1n})\vec{q} + k_2 (w_{21}, w_{22}, \cdots, w_{2n})\vec{q} \\ &= k_1 \rho(d_1, q) + k_2 \rho(d_2, q) \end{aligned}$$

As $0 < k_i \leq 1, i = 1,2$,

$$\rho(d_1 + d_2, q) \leq \rho(d_1, q) + \rho(d_2, q)$$

The mapping ρ is sublinear for document set **D** under TFIDF indexing. The main cause of ρ 's sublinearity is $\max_i Freq_{ij}$, which is usually regarded as normalization factor.

3. DISCUSSION AND IMPLICATION

Mapping ρ defined in the paper is sublinear for document set **D** under TFIDF indexing framework. The main element that matters is $\max_i Freq_{ij}$.

Firstly, we need to discuss the condition on which ρ is linear for **D**. Obviously in the inequity

$$\rho(d_1 + d_2, q) \leq \rho(d_1, q) + \rho(d_2, q)$$

the equal sign stands only when the equation

$$\max_{1} Freq_{1j} = \max_{2} Freq_{2j} = \max_{1+2} Freq_{(1+2)j}$$

is satisfied.

Although the equality above needs to be met only to keep the bilinearity of VSM, it is very strict in fact. It requires that on the dimension of the most frequent terms, the two documents in consideration must be orthogonal. This means the content of the two documents must be totally different on the two dimensions. This is very strict in real situation. $\max_i Freq_{ij}$ is the cause of variation of vector space geometrically. From information retrieval aspect, the variation is caused by indexing and weighting process. More precisely, the loss of similarity is large for similar topics, while it is small for different topics.

Secondly, we will discuss on $\max_i Freq_{ij}$. It is usually regarded as normalization factor. Discussions on normalization factor have been fruitful in past researches. Sparck Jones, Walker, and Robertson deemed that normalization factor was related to document length and average document length [8, 9]. Singhal, Buckley, and Mitra experiment explained that there was a cross between the probability curves of relevance and retrieval [7]. When using document length to normalize, the linear combination of previous normalization method and pivot. The paper explicates the normalization factor's effect on the vector space. If we do not take the normalization factors into consideration, i.e., taking raw frequencies as TF part in term indexing, indexing will not change the linear property of VSM's inner product space. This means the normalization factors cause the variation of linear characters of vector spaces. The influence of different indexing method is heterogeneous and needs discussions separately.

4. CONCLUSION AND FUTURE WORK

The discussion in the paper explicates that vector space model is a formal framework and that it is affected greatly by indexing method. Strictly speaking from algebraic deduction, it will not fulfill the requirements for bilinear functions under TFIDF indexing scheme selected in the paper. The main factor that accounts is normalization factor. In TFIDF indexing scheme, the maximum frequency of one document is of great meanings geometrically and for information retrieval.

"Bag of words" is the assumption of the paper, which guarantees the expansion of $\alpha(d_1 + d_2)$. With the development of n-gram indexing, the "bag of words" assumption might be replaced. While n-gram indexing hitherto is not good enough, the "bag of words" assumption is still widely accepted in the present IR research. The paper thus contributes to Vector Space Model under "bag of words" assumption only.

These results are still very elementary. We have discussed on the basic situation of Vector Space Model. Related issues such as clustering and relevance feedback are not covered in the paper, which leaves many problems to be resolved. Discussion on maximum term frequency and normalization factors need further study on its role in information retrieval and geometry.

5. ACKNOWLEDGMENTS

The author is partially supported by China Scholarship Council (No.: 2009601175) and Graduate School of Peking University.

6. REFERENCES

[1] Cater, S. C. 1986. The Topological Information Retrieval System and the Topological Paradigm: a Unification of the Major Modals of Information Retrieval. Doctoral Thesis. Louisiana State University, Baton Rouge.

[2] Dominich, S. 2000a. Foundation of information retrieval. Mathematica Pannonica, 11, 1, 137 – 153.

[3] Dominich, S. 2000b. A unified mathematical definition of classical information retrieval. Journal of the American Society for Information Science, 51, 7, 614 – 625.

[4] Dominich, S., & Kiezer, T. 2007. A measure theoretic approach to information retrieval. Journal of the American Society for Information Science and Technology, 58, 8, 1108 – 1122.

[5] Egghe, L., & Rousseau, R. 1998a. Topological aspect of information retrieval. Journal of the American Society for Information and Technology, 49, 13, 1144 – 1160.

[6] Egghe, L., & Rousseau, R. 1998b. A theoretical study of recall and precision using a topological approach to information retrieval. Information Processing and Management, 34, 2/3, 191 – 218.

[7] Singhal, A., Buckley, C., & Mitra, M. 1996. Pivoted document length normalization. In Frei, H., Harman, D., Schäuble, P., & Wilkinson, R. Eds. Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (Zurich, Switzerland, August 18-22, 1996). SIGIR'96, ACM Press. New York, NY, 21 – 29.

[8] Sparck Jones, K., Walker, S., & Robertson, S. E. 2000a. A probabilistic model of information retrieval: development and comparative experiments, part 1. Information Processing and Management, 36, 6, 779 – 808.

[9] Sparck Jones, K., Walker, S., & Robertson, S. E. 2000b. A probabilistic model of information retrieval: development and comparative experiments, part 2. Information Processing and Management, 36, 6, 809 – 840.

[10] van Rijsbergen, C. J. 2004. The Geometry of Information Retrieval. Cambridge University Press, Cambridge (Eng.)

Beyond Intent: Technology Adoption and Appropriation by University Staff

Pablo-Alejandro Quinones
University of Michigan
1085 Beal Ave
Ann Arbor, MI 48109
pabloq@umich.edu

Stephanie Teasley
University of Michigan
1085 Beal Ave
Ann Arbor, MI 48109
steasley@umich.edu

ABSTRACT
In this paper, we propose a model for understanding adoption and appropriation of technology. We describe a university-wide system that is designed for faculty and students, but which has been adopted by staff, followed by a survey study and some preliminary results.

Categories and Subject Descriptors
H.4.0 [**General**]

General Terms
Management, Measurement, Experimentation, Theory.

Keywords
Adoption, Appropriation, Courseware

1. INTRODUCTION
This paper reports on a work in progress investigating how a learning management system (LMS) designed for faculty and students that has been adopted by university staff. LMS log data show that university staff are using the "project site" capability of this system, which leads us to ask a series of research questions: Which staff are using the system? What do they use it for? Why do they choose to use it (or not)? Do staff use it in standard or innovative ways? We draw on the literature on technology adoption and appropriation to propose a model to frame our thinking about adoption and use of this particular system and other technologies.

2. BACKGROUND AND FRAMEWORK
Researchers have been interested in technology adoption and diffusion issues in organizations to predict a technology's success. If the new technology or tool is not incorporated into the existing workflow in meaningful ways, it is more likely to fail. For example, the success of a groupware application can be tied to its successful adoption by collaborators in an organization or work group [3]. One issue with adoption studies is that they tend to consider technology as static entities, that is, the technologies do not change in terms of their role and purpose. They are inserted into a group or organization (which is also usually static) and the technologies are used out of the box without modification with placid compliance to the designers' intentions. Researchers have shown that once it is released, technology is not static; it is often reconfigured and redefined by its users. Several researchers have used different terms to talk about this, but the term I will use for this is 'appropriation,' after DeSanctis and Pool [2] and Orlikowski [4].

There are very few models of appropriation presented in the literature and fewer that consider the re-design process. As one of main contributions, we present a model of technology appropriation adapted from Carroll [1]. We treat appropriation as one of four outcomes of evaluation that also includes disappropriation (abandonment), non-adoption, and simple adoption. This model avoids a flaw of adoption studies, where outcomes are only either adoption or non-adoption. It allows us to think of appropriation as a qualitatively different outcome than adoption. This particular treatment of adoption and appropriation distinguishes between adoption and appropriation as discrete outcomes for the sake of simplicity, though in actuality, they are the extremes out possible outcomes. We acknowledge that flawless adoption and complete appropriation rarely occur and that most adoption outcomes lie somewhere in the middle.

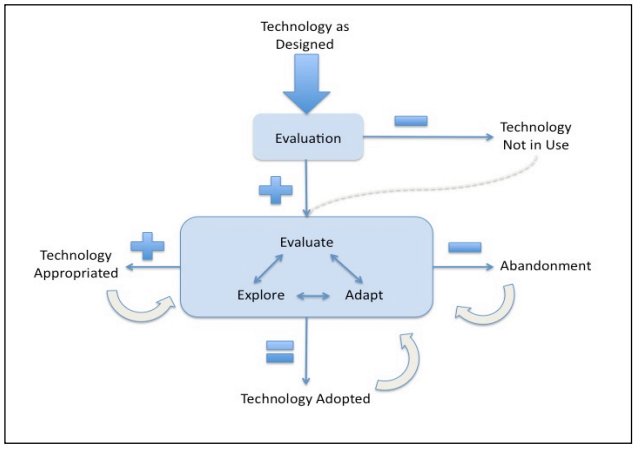


Figure 1: Model of Adoption and Appropriation

3. PROJECT SITES

To evaluate our model of appropriation, we are analyzing data collected as part of a university-wide study of LMD use. The LMS used on our campus include the ability to create project sites as part of the larger suite of courseware tools designed for faculty and students. This system, based on the Sakai architecture (see www.skaiproject.org), supports coursework and learning like similar systems such as Blackboard, Moodle, and ANGEL. Course sites are designed to support 'blended learning' [cite?] where the tools are used to supplement traditional face-to-face classroom interaction. Only course instructors can create these sites that are then automatically populated by the student lists provided by the registrar's office. Project sites include the same set of tools as course sites and act, look, and feel the same as course sites and are accessible from the same interface and menu as course sites. Unlike course sites, however, anyone on the campus can create a project sites including students, faculty, and staff.

For the purpose of design intent, is important to note that the only difference between course sites and project site is the label and who can create the site; the tools available on both type of sites are virtually are the same. This suggests that, for project sites, the design intent is to support the same kinds of course-related activities that students and faculty are engaged in. That students and faculty use project sites is not surprising since they already spend time in the course sites and become very familiar with the tool set. However, preliminary analysis of the site creation activity shows a surprising number of staff are setting up sites, surpassing even faculty site creation although not student's (see Table 1). Although it might be assumed that staff are setting up these sites for faculty or students, Further investigation reveals that staff members are creating sites primarily for themselves for administrative purposes (see Table 2).

Table 1: Project Site Creation

	2008-2009
Staff	23%
Faculty	18%
Students	59%

Table 2: Reported Site Purpose (as specified at creation)

	Learning	Research	Admin	Personal	Group	Other
Staff	14%	15%	54%	2%	3%	12%
Fac	32%	38%	24%	3%	4%	8%
Stud	29%	19%	2%	5%	38%	7%

4. METHODS

We administered an online survey at the University of Michigan over a three week period during the summer of 2009. We used a branched–survey design, where the answers to specific questions directed the respondents to different sections of the survey. The first part, which everyone completed, asked about general IT use. The last question in this section asked users to tell us about their use of Project Sites. From here, there were five branches: 1) those who have never used Project Sites, 2) those who have logged on once, 3) those who have logged in a few times, 4) those who were past users but currently did not use Project sites, and 5) those who are current project site users. Current users were then asked questions about their activity with the tool.

We invited 29,370 staff members to complete the survey. Of these, 4,672 staff members responded, for a response rate of nearly 16%.

5. PRELIMINARY RESULTS

Our branching question identified the extent to which the survey respondents were familiar with project sites. For respondents who had never used project sites, we asked them whether or not they had ever heard of this system. This allowed us to differentiate the differences between respondents who chose not to ever try the system from those who simply didn't know project sites exist. The results are shown in Table 3. The results suggest that about an equal number of staff who try project sites reject it (27%) as those who make use of it, whether in ongoing activities or for short durations (28%).

We asked our users to respond to the value of Project Sites for certain kinds of job activities on a 5-point Likert scale (1=Strongly Disagree and 5=Strongly Agree). Results shown on Table 4. The results suggest that communication, distance work, and providing a single access point for information were the biggest benefits.

We also asked all participants who had used the system at least once but were not currently users to react to various possible reasons why they discontinued their use. Ratings are on a 5-point Likert scale (1=Strongly Disagree and 5=Strongly Agree). Table 5 shows that the primary reason staff with only one login never used Project Sites was because they were "just looking." For those staff who had some limited experience with Project Site, the top reason for discontinuing use was because the specific project ended. For all respondents who were not current users, the fact that they had no co-workers using the system or that it had little connection to their job were the next most prevalent reasons for discontinuing use.

Table 3: Experience with Project Sites

Never used / Never heard of system	15%	690
Never used / Heard of system	30%	1389
One Login	7%	317
Few Logins	20%	898
Past User	10%	442
Current User	18%	840

Table 4: Mean Value-ratings for Project Site Features

Scheduling	2.98	1.031
Communication	3.87	.943
Posting Audio/Visual Materials	3.56	.928
Posting Group Materials	3.66	.947
Single Access Point	4.53	.679
Creating Groups	3.82	.904
Tracking Progress	3.47	.932
Distance Work	3.98	.862

Table 5: Top Three Reasons for ending Project Site Use

Activity	Reason	Mean	S.D.
One Login	Just looking	3.57	1.067
	No co-workers using	3.52	1.032
	Little connection to job	3.37	.898
Few Logins	Project Ended	3.56	1.048
	No co-workers using	3.37	1.056
	Little connection to job	3.17	.997
Past User	Project Ended	3.73	1.186
	No co-workers using	3.24	1.225
	Little connection to job	2.85	1.086

6. DISCUSSION

Thus far, the results have confirmed appropriation by some staff. Preliminary findings show that many university staff have adapted project sites to their work. Of those who try the tool, it seems that the biggest reason for non-adoption and abandonment is the lack of a project that requires it. This suggests that although these people have abandoned the tool, that they might be willing to use it again if they had a new project that required its use. Our analysis also suggests that there are two main uses for the tool—for enabling remote work and for a single location for various group materials.

Future analyses of our data will look to model which university staff have become users of the technology, as well as the extent to which these staff have adopted or appropriated the system. We will investigate questions addressing whether staff adapt project sites more broadly into their work than just for specific projects and show evidence of how these users have appropriated the technology to their work.

7. REFERENCES

[1] Carroll, J. (2004). "Completing design in use: closing the appropriation cycle", Proceedings of the 12th European Conference on Information Systems (ECIS 2004), Turku, Finland, 11pages.Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[2] De Sanctis, G. and Poole, M.S. (1994). Capturing the complexity in advanced technology use: adaptive structuration theory, Organization Science, 5(2), 121-147.

[3] Mark, G & Poltrock, S (2001). Diffusion of a Collaborative Technology Across Distance, *Group 2001 Proceedings*, 232 – 241.

[4] Orlikowski, W.J. (2000b). Using technology and constituting structures: a practice lens for studying technology in organizations. Organization Science, 11(4), 404-428.

Extending an LIS Data Curation Curriculum to the Humanities: Selected Activities and Observations

Allen H. Renear, Molly Dolan, Kevin Trainor, Trevor Muñoz
Center for Informatics Research in Science and Scholarship
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 East Daniel Street
Champaign, IL 61820
1-217-333-3280
{renear, madolan, trainor1, munoz14}@illinois.edu

ABSTRACT
We describe selected activities of an IMLS-funded project to extend an existing data curation curriculum to include humanities data, making a number of preliminary observations and conjectures.

Categories and Subject Descriptors
K.3.2 [**Computing Milieux**]: Computer and Information Science Education – *curriculum, information systems education*

General Terms
Management, Documentation, Reliability, Standardization.

Keywords
Data curation, humanities, education, curriculum, libraries

1. BACKGROUND
Although originally focused on data in the natural sciences, the data curation community is now concerned with data in the humanities as well. To help prepare the next generation of information professionals, the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign received a grant from the Institute of Museum and Library Services (IMLS) to extend our existing Data Curation Education Program (DCEP) [1], which is also IMLS funded, to include humanities data. This new program, DCEP-H, began in August 2008.[6] Among the activities underway are needs analysis studies, curriculum design, case study development, a fellowship program, an internship program, a summer institute for in-service training, and the development of a framework of cross-disciplinary data curation concepts. Through the IDEA group DCEP-H collaborates with other IMLS-funded data curation education programs to share insights and experiences. [3].

2. SELECTED ACTIVITIES
In this abstract we discuss three DCEP-H activities that are contributing to improving our understanding of humanities data curation needs. For general information about the project and about other DCEP-H activities see Renear et al. [6]

2.1 The 2009 Summer Institute for Data Curation in the Humanities
The DCEP 2009 Summer Institute for Data Curation, held May 18-21, 2009, at the Graduate School of Library and Information Science, Urbana-Champaign, Illinois, focused on humanities data and was coordinated by DCEP-H and funded by IMLS. The Institute brought together subject specialists, archivists, digital library, metadata, and repository professionals, as well as those who work directly with research data, for a discussion of emerging themes and practices in the management and preservation of humanities data.

2.1.1 Sessions and Instructors
A number of sessions were presented, including:

- *Curation of research data: Understanding scholarly practices and the role of collections,* **Carole Palmer & Melissa Cragin (Illinois).**
- *Descriptive markup and the TEI,* **Julia Flanders & Syd Bauman (Brown University).**
- *Markup semantics and the preservation of intellectual content: the data curator reads Ecclesiastes,* **Michael Sperberg-McQueen (MIT/W3C).**
- *Tools for Textual Data,* **John Unsworth (Illinois).**
- *Digital Preservation and Institutional Repositories,* **Dorothea Salo (University of Wisconsin).**
- *An appropriate institutional response?: Some organizational contexts for digital asset management,* **Lorcan Dempsey (OCLC).**

2.1.2 Attendees
The 24 attendees of the Summer Institute came from a wide variety of institutions and roles. Of the librarians that attended seven were subject or area specialists (e.g., Special Collections, Visual Resources, Arts, English & Theatre), five were cataloguing or metadata librarians, five worked with digital content, and two were data librarians. There were three attendees from digital humanities as well as a NOAA data librarian and a content publisher.

2.1.3 What we learned
Several things stand out:

1) **Demand**: We received 43 applications, almost all highly qualified, and were only able to accept 24. All 24 of those accepted chose to attend.

2) **Satisfaction**: We were surprised at the degree of excitement and satisfaction expressed by our attendees. This is certainly another sign of unmet need.
3) **Interest in additional formats**: Interest was repeatedly expressed in additional in-service education, adapted to specific roles (managers, curators, technical staff), and at different levels.
4) **Career level**: Most of our applicants were mid-career professionals: (e.g., Data Services Librarian, Metadata Librarian, Visual Resources Librarian, Special Collections Librarian.) They were clearly preparing for anticipated management of new curatorial activities.

The demand for in-service education in humanities data curation is very high. Interest appears especially strong from managers acquiring new responsibilities, and programs focusing on those needs are in order. Valuable programs for those taking on these new roles can be developed on the model of the Summer Institute. However there is also need for a variety of formats serving different target audiences.

2.2 Towards Common Concepts

A common framework of curatorial concepts and terminology that can be applied across domains would have many advantages for data curation.

Unnecessary diversity in terminology creates a burden for design, implementation, documentation, and training. A proliferation of discipline-specific frameworks can also impede the recognition of opportunities for simplification, efficiency, and the re-use of successful strategies across domains.

Common frameworks of concepts could help (i) simplify and unify policies and documentation that support curatorial activities, and (ii) support a more uniform data curation curriculum. But can such distinct domains of scholarship as the sciences and humanities support shared frameworks of common concepts?

One important notion that appears to exist in many different domains is the distinction between data that is relatively raw or in some sense accepted as given, and data that is "processed" or the result of interpretation and analysis.

As an exercise in conceptual alignment in this area, we have compared the widely used NASA Earth Observing System data level categories with traditional notions of levels of editorial intervention found in textual criticism.[5] The relative ease of finding intuitive alignments across domains provides some evidence that a shared framework in this area is possible. However an obstacle to a confident alignment was that neither NASA nor the literature of textual philology provides an adequate conceptual (as opposed to merely operational) account of what data levels are, focusing instead on what features should be at what level. We suspect this will be a problem for other alignment projects as well.

Given these encouraging results we are beginning a collaboration with the Data Practices and Data Concepts groups of the NSF-funded Data Conservancy.[4] These groups, which are also based at GSLIS, are among other things developing a formal framework of cross-cutting fundamental concepts related to data curation in the sciences.

2.3 Assessing the Skills Needed for Humanities Data Curation

To determine what skills are needed by humanities data curators we have two projects that focus on the workplace:

2.3.2 Analysis of Position Descriptions

DCEP-H is working with an ongoing DCEP project (led by Melissa Cragin) to assess job postings in the sciences, social sciences, and humanities for skills and education relevant to data curation.[2] Job announcements were collected from a number of online job hosting sites. Relevancy was determined by the collector via a two-stage process of (i) assigning general relevancy by identifying data curation terminology and (ii) assigning specific relevancy by manually examining the announcement in detail. Analysis of the collected data will begin in January 2010.

2.3.3 Survey/Interview Needs Assessment

For another perspective on the skills needed for curation of humanities data we are conducting a survey and structured interviews with management and professional staff at selected humanities computing centers to identify problems faced in data management, as well as current best practices and future needs for data expertise.

3 ACKNOWLEDGMENTS

This program is sponsored by the Institute of Museum and Library Services (RE-05-08-0062-08).

4 REFERENCES

[1] Cragin, M.H., Heidorn, P.B., Palmer, C.L., Smith, L.C. (2007). An Educational Program on Data Curation. Poster, Science and Technology Section of the annual American Library Association conference. Washington, D.C., June 25, 2007. http://hdl.handle.net/2142/3493.

[2] Cragin, M., Palmer, C.L., Varvel, V., Collie, A., Dolan, M., (2009, December). Analyzing Data Curation Job Descriptions. Poster, 5th International Digital Curation Conference, London, England, December 2-4, 2009. https://www.ideals.illinois.edu/handle/2142/14544.

[3] Hank, C., Davidson, J. (2009). International Data curation Education Action (IDEA) Working Group: A Report from the Second Workshop of the IDEA. *D-Lib*, 15(3/4). http://www.dlib.org/dlib/march09/hank/03hank.html.

[4] Palmer, C. L. (2009, December). The Data Conservancy: A Digital Resource & Curation Virtual Organisation. Presentation, 5th International Digital Curation Conference, London, England, December 2-4, 2009.

[5] Renear, A.H., Dolan, M., Trainor, K., Cragin, M.H. (2009). Towards a Cross-Disciplinary Notion of Data Level in Data Curation. Proceedings of the 72nd ASIS&T Annual Meeting. Vancouver, BC, November 8-11, 2009. http://www.ideals.illinois.edu/handle/2142/14547.

[6] Renear, A.H., Teffeau. L.C., Hswe, P., Dolan, M., Palmer, C.L., Cragin, M.H., & Unsworth, J.M. (2009). Extending an LIS Data Curation Curriculum to Include Humanities Data. Poster, DigCCurr 2009 conference. Chapel Hill, NC, April 1-3, 2009. https://www.ideals.illinois.edu/handle/2142/14548.

Peer Production in Politics: Democracy vs. Governance in the Fifth Estate

Jessica Richman
Oxford Internet Institute
1 St. Giles
Oxford UK 3 OX1 3JS,
+44 1865 287 210
jessica.richman@gmail.com

ABSTRACT

As Andrew Chadwick noted in 2006, "The issue is no longer whether politics is online, but in what forms and with what consequences?"[1] Governments use the Internet for many purposes, including informing the public, facilitating voting, and communicating with constituents. Similarly, citizens engage with political institutions online in a variety of ways. William Dutton, who terms this engagement of citizenry with government the "Fifth Estate", delineates two institutional arenas: the use of "the Internet and related ICTs to enhance existing democratic institutions and processes…and the networking of individuals to enable the public to hold all institutions of government and politics more accountable". [2] This poster will contrast websites that use the Internet to facilitate representation (e.g., the Sunlight Foundation, Wikileaks) with those that focus on governance (e.g., Community Patent Review, SeeClickFix).

These organizations differ in their fundamental objectives: the first type uses the Internet to influence democracy itself while the second engages citizens to do work already delegated to government. There are many examples of both types of organizations, both in the United States and abroad. [3] Some use crowdsourcing to create or organize information; others present information to users but do not allow them to contribute. [4] This poster will focus on organizations that operate in the US and use peer production; it will highlight their use of online collaboration, as well as their impact on issues of government transparency, democratic representation, and peer production of knowledge.

The mission of the Sunlight Foundation is to use "the revolutionary power of the Internet to make information about Congress and the federal government more meaningfully accessible to citizens."[5] As such, it sponsors online tools that allow the public to watch and contribute to knowledge about government spending, the legislative process, and the influence of lobbyists. One project of the Sunlight Foundation, PublicMarkup.org, allows the public to annotate bills before they are passed by Congress. Several 2008 bills were posted on the site, including the act that authorized the Troubled Assets Relief Program (TARP). The Sunlight Foundation also sponsors projects that do not use peer production, such as MAPLight.org, which allows combines data on campaign contributions and voting records for legislators.

Wikileaks.org allows users to upload "leaks" -- information about corrupt officials, scandals, and other misbehavior – in an anonymous manner. Wikileaks thus crowdsources the whistle-blower role in government (as well as other areas) and also the fact-checking of whistles blown. Wikileaks operates in nearly every country (as it is a wiki, countries can be added very easily), but is represented by Australian Hacker Julian Assange. [6] Situated between journalism and governance, Wikileaks is more difficult to classify, but as an example of whistle-blowing on government corruption, it aims to improve the democratic process.

Community Patent Review is a striking example of an organization that uses peer production for governance. Community Patent Review is a "web-based application that allows third parties to comment on patents before they are issued. Large patent holders, including General Electric, Hewlett-Packard, IBM, Intel, Microsoft, Oracle, and Red Hat, have agreed to be part of the pilot, and the patent office has agreed to waive the usual fee for third party comment submissions and give the patents of pilot participants expedited review."[7] This provides a public adjunct to the Constitutionally designated function of the Patent Office; this citizen participation does not follow the chain-of-command of the patent office, but instead involves citizens (and non-citizens) in a new way, as users and collaborators.

SeeClickFix operates at the local level. SeeClickFix "allows anyone to report and track non-emergency issues anywhere in the world via the Internet. This empowers citizens, community groups, media organizations and governments to take care of and improve their neighborhoods." [8] Users post local maintenance issues (e.g., potholes, vandalism), and others can volunteer to fix them. This enhances the role of existing local government functions (e.g., the parks department, the roads department) by again involving the public directly in a non-democratic capacity.

Each of these organizations uses a different method to apply peer production to the task of improving democracy or of improving governance. Sometimes these tasks are mixed: MySociety, a UK organization similar to the Sunlight Foundation, creators of FixMyStreet (which inspired SeeClickFix in the US), hosts an e-petition process, which facilitates, instead of street fixing, the development of petitions that are sent directly to the Prime

Minister.

Organizations can be difficult to classify by these criteria. Creative Commons, founded in 2001, uses open licensing to bypass the more restrictive features of copyright law. Although it is not technically "crowdsourced", it does rely on the network effects of its many users who can build on each others' creations as long as they are all licensed under Creative Commons. In this aspect, it is an example of the second type of organization. Instead of lobbying to change copyright law, it bypasses the structure chosen by democratically elected representatives and instead uses a "Fifth Estate", user-governance model. People decide for themselves what kind of copyright they would like to have and share that with others. Similarly, SeeClickFix does not try to recall local parks officials, but instead points out where citizens can do local maintenance themselves; Community Patent Review does not aim to reform the US Patent Office, but instead facilitates public input.

The subject of the first group is representation; the subject of the second is a social issue, such as property rights in innovation or local property damage. Both approaches involve transparency, but the second approach involves further transparency as the citizens are full participants in the entire process. For example, what a government does with Wikileaks information or how the annotations on laws are processed is not necessarily transparent. Whether or not a problem has been fixed on SeeClickFix is transparent. In other words, the first category still involves representation, which naturally decreases transparency, while the second does not.

These two approaches are in some degree of tension: if, in the extreme case, all government functions were handled directly by citizens, there would be no need for representative democracy. On the other hand, if transparency were all that democracy needs, then perhaps the more communitarian approach would not be necessary. In some ways, these two approaches are, to use an economic term, substitutable goods.

Both of these approaches are simply the result of applying peer production to public goods, such as well-maintained public spaces, efficient patent issuance, or well-formulated laws. As these organizations grow, they will necessarily blur these lines and create further emergent forms of online democracy and online governance.

This distinction is significant because as peer production becomes more common and as both governance and the political process use information technology to an increasing degree, these tensions will come to the forefront. This is important along two dimensions: from an academic perspective, our understanding of the meaning of democracy (including representative democracy) and the position of the relevant actors will be in flux. From a practitioner perspective, an understanding of these differences can lead to additional opportunities to use peer production to improve governance. In addition, the convergence of these methods creates hybrid forms that will require a deeper understanding of the relation not only of citizens to the state, but of the nature of representation and collective action in general. In summary, the "Fifth Estate" may prove to be as important to both the conception and practice of governance as the famed Fourth Estate, journalism.

Categories and Subject Descriptors

K.4.3 [**COMPUTING MILLEUX**]: Organizational impacts – *computer-supported collaborative work*. K.4.1 [**COMPUTING MILLEUX**]: Public policy – Regulation.

General Terms

Management, Measurement, Design, Economics, Human Factors, Standardization, Legal Aspects.

Keywords

Politics, Web. 2.0, governance, user-interface design, online communities, public policy, Internet, ICTs

1. ACKNOWLEDGMENTS

I would like to acknowledge Professor William Dutton and the Oxford Internet Institute. In addition, I am grateful for the support of the Clarendon Fund.

2. REFERENCES

[1] Chadwick, A. 2006. Internet Politics, Oxford University Press.

[2] Dutton, W. 2006. Prometheus, Volume 27, Number 1, March 2009, pp. 1-15(15)

[3] For example, MySociety.org in the UK or online legislative deliberation in Brazil.

[4] For example, FiveThirtyEight.com, which presents election information in the United States, or MapLight.org, another project of Sunlight Labs, which combines campaign contribution information with legislative voting records.

[5] http://www.sunlightfoundation.com/about/ Accessed November 14, 2009.

[6] http://www.p2pnet.net/story/16318 Accessed November 17, 2009.

[7] Richman, J. 2008. UC Berkeley Science, Technology and Engineering Policy White Paper Competition. http://step.berkeley.edu/White_Paper/richman.pdf

[8] http://www.seeclickfix.com/how_seeclickfix_works Accessed November 17, 2009.

Technology Access and Training in Public Libraries: A pilot study of technology assistance to patrons of the Urbana Free Library

Susan Rodgers
Graduate School of Library and Information Science, University of Illinois Urbana-Champaign
501 E. Daniels Street
Champaign, IL
1 (217) 333-7094
srrodger@illinois.edu

ABSTRACT
In this extended abstract, an ongoing pilot study looking at technology assistance in the Urbana Free Library computing lab is described. The study is student driven and is an example of how a research project can provide quality service to the community in which it functions.

Categories and Subject Descriptors
K.3.2 [**Computers and Education**]: Computer and Information Science Education – *literacy*

General Terms
Management, Measurement, Theory.

Keywords
Participatory Action Research, Community Informatics, Technology Education, Information Literacy, Sustainability, Pedagogy

Introduction and Background
In the spring of 2009 two graduate students from the University of Illinois Urbana-Champaign set out to organize a volunteer technology force at the local public library. A group that now numbers over 30 volunteers was developed with a goal to research the types of support provided to patrons using the public computing lab. The volunteers are now in the process of conducting a pilot study on what types of questions and technology needs are being asked in the library computer lab. More specifically, the research examines the role that public libraries play in providing information literacy training to patrons who use their technology services with an aim to find out what the information literacy needs are that libraries are addressing. The findings from this research will be a starting point from which to explore different models of information literacy training.

There have been a few major research studies on the technology services that libraries provide, notably, Bertot's u2007 study, which sampled 6,979 public libraries in metropolitan, suburban and rural areas around the U.S. The study found that while almost all libraries provided free Internet access to their communities, and 73% were the only provider of free Internet access in their service areas (77% in rural areas), only 30% provided computer skills and Internet training. While the most commonly reported outcome of the training was "Provides information literacy skills (46%)," it is unclear what the nature of the training (formal group training, informal one-on-one training, or scheduled one-on-one training) is.

The Bertot study gives us a comprehensive picture of the technology-related services that U.S. libraries are providing, and a glimpse into what the outcomes are, but little clue as to what patrons are learning, how that learning is happening between patrons and librarians or the impact this has on the job of the librarian. Bertot suggests future research is needed on expectations of patrons, community and governments, specifically, what users want from their public library's Internet access and training services.

Research Focus
This study aims to find out what technology-related inquiries people are asking the library for help with and how much time it takes to resolve these queries by examining interactions between patrons and volunteers who provide assistance with computer and Internet-related questions at the adult reference desk at The Urbana Free Library (UFL) in Urbana, IL. By collecting data on these interactions, the researcher will identify user needs, learn more about how these needs are affecting the job role of the librarians, and suggest different models for information literacy instruction that will meet the needs of users.

This research will add to a body of knowledge about the kinds of formal and informal training that is happening in public libraries. It will also give a clearer picture about the extent to which public libraries are providing staff time and support for information literacy training. As libraries are often the only provider of free Internet access in a community, this information will be helpful for libraries and local government in terms of planning and staffing for technology-related services. It should be noted that this is a pilot study of just one method of delivering information literacy training. Additional research will be needed to adequately

address the question of which training models will fulfill patron needs.

Site Location

The Urbana Free Library is a public library funded through city taxes. Its policies are determined by a library board appointed by the mayor. Computer lab user cards for logging in are available to any individual, no matter their residence, in addition to "guest passes" for one day use. Patrons are allowed two computer logins a day at 30 minutes per session. However, time extension requests are rarely denied. The lab contains 29 computers broken down into the following categories: 6 non-Internet workstations, 2 express Internet workstations, 19 full-service workstations, and 1 full-service Internet workstation with a large monitor and assistive technology equipment. The lab is available at all times during library hours and is monitored by the adjacent adult reference desk.

Approximately 37,000 people live in Urbana, Illinois. The population is 14% African-American, which is higher than the national percentage (12%). 13% of families and 27% of individuals live below the poverty level (Census, 2000). In addition to a large African-American population and groups below the poverty level, nine months out of the year 40,000 students attend the University of Illinois located in both the cities of Urbana and its adjacent city Champaign (On-Campus Enrollment, 2009). Students are a significant portion of the population during the school year and many of them are patrons of the Urbana Free Library.

The Urbana Free Library was an ideal choice for this study because of an existing relationship between the staff at the library and the primary researcher on this project. The librarians made it known that they are in need of assistance in the computer lab, and that they trust the researcher with the authority to organize volunteers, partly because the researcher is affiliated with the University of Illinois Graduate School of Library and Information Science (GSLIS). They are also willing to allow the researchers to collect data from interactions with patrons in the lab. Most of all, many of the reference librarians made it clear that there is an urgent need for volunteer assistance as patron demand for technology-related assistance is making it difficult for reference librarians to balance these duties with the more traditional duties of reference, readers' advisory, collection development and other tasks.

Method

To gather data for this study the majority of volunteers are solicited through the Community Informatics Club, a GSLIS student organization. The volunteers attend an initial training session where they were familiarized with the computer lab, the procedures for working in the lab, and are introduced to the research and data collection in which they play a key function. Since this volunteer group was developed last spring a number of other individuals, mostly graduate students at the University of Illinois, have stepped in to provide support in the lab.

There are three priority volunteer days per week: Tuesdays from 3-7 p.m., Thursdays from 4-8 p.m., and Saturdays from 1-5 p.m. Each volunteer day is broken into two-hour sessions with a one volunteer staffing the station at a time. These times were chosen so that volunteers would be available in the lab during the hours patrons are arriving from school and work in addition to a high-activity time over the weekend. In all, there is at least one volunteer in the lab for 12 hours a week. Additionally, the fall 2009 semester saw an influx of new volunteers and as a result, nearly each week additional shifts are added on in order to accommodate more helpers in the lab.

The method for data collection is a categorized statistics sheet that is filled out by the volunteers. For every interaction the volunteer has with a patron, they mark a tally in the category in which the question most closely relates such as *document creation, e-mail assistance, operating a computer, printing*, etc. The last field is an open-ended space that allows the volunteers to share additional information about the interaction. In this space, volunteers are invited to express difficulties, frustrations, or surprising results from their time with the patron.

Where the Project Stands

The research for this project is currently ongoing and still in the pilot stage. Initially, the purpose of this volunteer group was to primarily conduct the research and come to some end conclusion. However, the volunteers have been so successful in the assistance they provide, and the library so grateful to have the help, that the project has now moved its primary focus away from the data side of the work to a more comprehensive effort to build a sustainable volunteer group that will continue to provide much needed support to patrons in the lab. It is apparent that there is much more to learn here than what the researcher initially set out to accomplish and that what initially began as a research project has now developed into a greater effort with ramifications beyond the walls of the university and into the surrounding community. While still interested in the research, it is important that the volunteer group be further organized into a sustainable and permanent support team for the Urbana Free Library public computing lab.

References

[1] Bertot, J.C., McClure, C. and Jaeger, P. 2008 Public libraries and the Internet 2007: Issues, implications, and expectations. Library & Information Science Research 30, pp. 175–184. .

Understanding How Vulnerable Populations Use Common Information and Communications Technologies (ICTs) to Access Health Care Information

Michelle Rogers, PhD.
Drexel University
3141 Chestnut St.
Philadelphia, PA 19104
215.895.2922
mrogers@drexel.edu

Lisl Zach, PhD.
Drexel University
3141 Chestnut St.
Philadelphia, PA 19104
215.895.2476
lisl@drexel.edu

Prudence Dalrymple
Drexel University
3141 Chestnut St.
Philadelphia, PA 19104
215.895.2699
pdalrymple@drexel.edu

ABSTRACT

Overview

The goal of this study is to investigate patterns of internet access via computers and cellular telephones among the population being served by the Drexel University Eleventh (11th) Street Health Center (Here after, the Center) and to determine which Information and Communications Technologies (ICTs) will be most appropriate for delivering health care information to this population. This two-part study examines 2 concepts: 1). Which ICTs do patients at the11th Street center use to connect to the internet; 2) How these patients use the internet through the identified ICTs. Results of the project will be used to improve the ways the Center communicates with its patients by developing new approaches for using ICTs that better reflect the existing patterns of use in the population. The study will also provide researchers with a better understanding of the barriers and facilitators to internet access in this vulnerable population.

Background

Healthcare information technology (HIT) holds great promise for improving patient outcomes, increasing cost-effectiveness and both patient and staff satisfaction. However, implementation of HIT has been fraught with problems as organizations have struggled with the transition to e-health. Few settings experience this frustration more keenly than those that operate using a care delivery model that differs from the "normal" physician centered, medical model. One example of such a model is the nurse-managed, primary care clinic that coordinates patient services from a variety of health care professionals to deliver patient-centered care to populations that are medically under-served. Such nurse-managed care is likely to gain importance as the prevalence of chronic conditions requiring trans-disciplinary care increases in our society, especially among vulnerable populations, those who have little or no access to insurance-supported medical care.

To improve and manage the health status among the area residents, the center provides a wide variety of health and wellness services including physical exams, diagnosis and treatment of illness, family planning, health maintenance/disease prevention services, behavioral health services, physical fitness programs, dental services, nutrition services, and adolescent health initiatives. In particular, the center employs an innovative and trans-disciplinary care model that fully integrates behavioral and various wellness services into primary care to form a team approach.

Research Questions

- How do patients at the 11th Street Family Health Center access the internet?
- How do these patients use the internet?
- What barriers exist to internet access for these patients?

Setting

The Center serves an area of Philadelphia that the U.S. Department of Health & Human Services has designated as "medically underserved." The area population of 19,642 people is 90 percent African-American with a growing Latino (6%) population. The unemployment rate of 49 percent is the highest in the city. In addition, the median family income of less than $13,000 is the lowest in Philadelphia. With a focus on health education and chronic disease management, the Healthy Living programs of the Center help ensure that area residents can achieve their optimum state of long-term health and well being. The programs address priority community health concerns – diabetes, obesity, hypertension and prenatal care – while fostering social inclusion, reducing health inequalities, and increasing residents' use of health services and programs that complement city-wide and national strategies. The Healthy Living programs coincide with The Center's trans-disciplinary model of clinical care, through which health professionals cross-train in multiple specialties and create an integrated care team where knowledge and skills overlap. This allows a single provider to see a more complete picture of each patient and gives providers the opportunity to assess, and in some cases treat, patients in other health areas.

Methods

Initial data were collected in 53 face-to-face semi-structured interviews with patients at the Center (No identifiable data was collected). The interviews were tape-recorded for referential accuracy, and data are currently being analyzed for themes and patterns. In addition to the interviews, focus groups will be held to further explore key concepts identified from the interview data.

Preliminary Findings

Access:

- 72% of survey respondents reported having access to the internet, slightly lower than the 77% reported by the 2009 Pew Internet & American Life survey for total adult internet access in the general population. [http://pewinternet.org/Static-Pages/Trend-Data/Whos-Online.aspx <accessed 2 Nov 2009>]
- Results suggest that cell phones are used more frequently than computers to access the internet, especially by women 25 – 34.

Use:

- 45% of the sample report to use a computer several times a day
- Barriers to use include having no computer/internet connection; not knowing how to use one and being frustrated with use.

Searching for information:

- Email and recreational use of the internet were most frequently reported by computer users, while general information seeking was most common among cell phone users.
- Computer users are more likely to search for health information on the internet than those who use cell phones, but only 21% (8 of 38) of the patients reported using the internet for this purpose. The 2009 Pew study on The Social Life of Health Information reports that 61% of the general population shows similar use patterns. [http://www.pewinternet.org/Press-Releases/2009/The-Social-Life-of-Health-Information.aspx <accessed 2 Nov 2009>]
-

Next steps

- Identify patterns/trends in interview data to begin to inform the design of mobile interfaces as well as improve health information systems
- Examine potential to use cell phone technology to communicate health care information
- Investigate specific barriers and facilitators to seeking health information on the internet by this population
- Identify patterns/trends in interview data that support model of information-seeking behavior of patients at the Center

Researcher Subjects: Gaining Access and Building Trust in an Online Breast Cancer Support Group

Ellen L. Rubenstein
Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign
501 E. Daniel Street, Champaign, IL, 61820, USA
erubens3@illinois.edu

ABSTRACT
This poster presents an analysis of some of the challenges in gaining access and building trust in an online breast cancer support group. Although the group itself is not a closed group and is freely available on the Internet, divergent participant beliefs about privacy and public access to this website, as well as conceptions of research methods and ethical concerns, offer provocative insights into the perceptions online participants have about research, the role of researchers, and the juxtaposition of researchers and subjects within the context of this particular online community.

Keywords
Online Communities, Health Informatics, Online Ethnography

1. INTRODUCTION
Ethnographers have long described the challenges of gaining access for the purposes of conducting naturalistic research of social worlds. Communities are often suspicious or fearful of a researcher's motives and may initially be wary of engaging in interactions [1]. Prior membership in a group often confers greater legitimacy in the views of participants than approaching a group from the outside; however, gaining entry can also be facilitated through intermediaries with close ties to a community [2][5]. Familiarity with a community's norms and practices, or possessing attributes similar to community members, also facilitates acceptance [4]. However, even when a researcher can claim familiarity and similarity, questions and challenges can arise upon entry into a community.

The Breast Cancer Mailing List (the "List") is an online support group founded in the mid-1990s, comprised primarily of breast cancer patients and survivors, as well as advocates, researchers, family members, and medical professionals. Although the List requires a subscription to post messages, it is otherwise freely available to anyone wishing to read the archives. The List currently has between 400-500 subscribers, most of whom live in the United States, although there are participants from approximately 20 other countries.

With the intent to conduct research as a participant-observer, I approached the List to learn how health information is exchanged, how participation in the group influences health practices and outcomes, and how the group is integrated into their everyday lives. As a researcher with background knowledge of the group as well as familiarity with health information seeking and breast cancer survivorship, I anticipated minimal access problems; however, the reception was mixed. Gaining access to the group and eliciting trust comprised intensive interactive and iterative processes wherein wariness and suspicion were interspersed with expressions of welcome and hopefulness.

2. NEGOTIATING ACCESS
In accordance with IRB approval, my first message to the List revealed my status as a researcher, a Ph.D. student, and a breast cancer survivor. Responses included a range of reactions. Participants expressed concerns about my legitimacy, fear of how the research might be used, confusion about my role in conducting ethnographic research, and apprehension about use of quotations. Some participants expressed feelings of trepidation and vulnerability but were tentatively open to research on the group. Other participants were fully supportive, offering to assist in any way possible. To assuage concerns, it was essential to be totally open in providing information that would assure them that the research was legitimate and that I had no intentions of exploitation. I answered every question in great detail, sent out copies of my IRB form, and revealed personal information in the spirit of sharing as any other List participant would. In essence, the roles of researcher and subjects became intertwined as questions and evidence accrued.

2.1 Researcher Subjects
Online forums offer the potential for people to assume identities and deceive others for various reasons [4]. In online breast cancer support groups, participants are often dealing with serious issues in their lives, making them feel particularly vulnerable [3][6]. During my initial entry into the community, List members reported previous instances of marketers posing as researchers who tried to sell them products as well as researchers recruiting them for seemingly questionable studies. These experiences were viewed as both disruptive and intrusive, causing participants to seriously appraise newcomers identifying themselves as researchers. However, this process of appraisal illustrates the dual role in which I was cast: as a researcher I also became a subject of inquiry.

List members are highly educated and savvy, and due to their status as breast cancer patients and survivors, some have participated in clinical trials and most are very aware of trials occurring on an ongoing basis. Discussions about clinical trials as well as other health research occur frequently, often several times

a week. The constant reporting of medical research published in journals and news stories results in a List population that is sensitized to researcher and subject roles.

List participants' experiences with both legitimate and spurious research practices converged upon my appearance. Although there was recognition of the value of being research subjects, especially in clinical trials, ethnographic research was less familiar and elicited questions about my objectivity as well as my integration into the community. Participants became interested in verifying the genuineness of my research as well as the legitimacy of my methodology, and several did their own research to ascertain what ethnography constituted.

2.2 Privacy and Public Access

Within the context of my accessing the List, members expressed varying levels of understanding about their exposure on the Internet and who might be reading their words or observing their interactions. Some participants were fully aware that anything they wrote was completely accessible while others had not thought about the reality that their private expressions on the List were available to anyone. Some also expressed discomfort with the idea of a researcher using quotes from their discussions while others gave permission immediately. Throughout these interactions I assured participants that I would use quotations only with their express permission, subject to my IRB approval as well as my commitment to their privacy.

2.3 Building Trust

Throughout the access negotiation period, various List members expressed several attributes of valued participation. Most important, beyond establishing my credence as a researcher, was interest in my breast cancer journey. Participants began to feel more comfortable with me as my own story became known to them. Being an active participant was also essential to their overall acceptance of me. Although there was recognition that lurkers and infrequent participants were likely a substantive portion of List readers, as a researcher who was also a participant-observer, it was critical to be active. As their own stories unfolded, I offered responses when they were appropriate, both through acknowledging comments as well as descriptions of my personal experiences.

Another significant element to building trust was participation in the List's face-to-face Gathering. Despite the group's primary presence as an online resource available to anyone needing support with breast cancer and its aftermath, a major component to the List's interactions is its yearly fall meeting. The Gathering occurs somewhere in North America and is open to anyone who is a List participant, as well as family members and friends. Although a relatively small core of members attends, it is an opportunity to meet in person and solidify the bonds people have formed through their online interactions. Within days of my joining the List, several members invited me to attend as a way to meet me in person, assess who I was, and to become more comfortable with my presence. Through participation with List members both virtually and face-to-face, I was able to establish a foundation for a trust-based research relationship.

3. CONCLUSION

Gaining access to the List and building trust with participants comprised multiple components, all of which blurred researcher and subject roles. Throughout the process, I acted both as researcher and subject while List participants conducted their own inquiries about me. To gain trust and establish credibility as a researcher and a participant-observer, it was essential to openly answer all questions. It was also important to be an active participant, sharing personal experiences and engaging in dialog with List members. Last, meeting List members in person amplified their trust and acceptance of me.

4. REFERENCES

[1] Hammersley, M., & Atkinson, P. 2007. *Ethnography: Principles in practice* (3rd ed.). NewYork: Taylor & Francis.

[2] Fetterman, D. M. 1998. Ethnography. In L. Bickman and D. J. Rog (Eds.), *Handbook of applied social research methods*. Thousand Oaks, CA: Sage.

[3] Høybe, M. T., Johansen, C., & Tjørnhøj-Thomsen, T. 2005. Online interaction: Effects of storytelling in an Internet breast cancer support group. *Psycho-Oncology,* 14, 211-220.

[4] Kendall, L. 2002. *Hanging out in the virtual pub: Masculinities and relationships online*. Berkeley, CA: University of California Press.

[5] Lofland, J., & Lofland, L. H. 1984. Analyzing Social Settings: A Guide to Qualitative Observation and Analysis (2nd ed.). Belmont, CA: Wadsworth.

[6] Radin, P. 2006. "To me, it's my life": Medical communication, trust, and activism in cyberspace. *Social Science & Medicine*, 62, 591-601.

Collaborative modeling for robot design

Selma Sabanovic
School of Informatics and Computing
Indiana University
selmas@indiana.edu

Matthew Francisco
Department of Science and Technology Studies
Rensselaer Polytechnic Institute
francm@rpi.edu

In this poster, we describe a method for using grounded theory and modeling to support collaborative design of social robots for the elderly. Robotic technologies are being designed to assist people in their everyday lives in various ways: as companions [9], domestic helpers [4], receptionists [1], and educational aids [8]. In response to the steadily rising average age of the population in the US, Europe, and Japan, the elderly are often designated as an appropriate audience for assistive robotic technologies. Designing robots for the elderly poses a variety of social challenges—understanding the specific needs and desires of the elderly, supporting independence and human dignity, and making sure that technologies can be successfully incorporated into existing social and physical environments, or "elder ecologies" [3]. These challenges suggest that designing robots for the elderly calls for attention to individual attitudes towards technologies as well as community norms and practices of social interaction and technology use.

1. DESIGNING WITH THE ELDERLY

In designing robots that can participate in the daily lives of the elderly, we propose working *with* rather than just *for* the elderly. The elderly are a vulnerable population whose worldviews and expectations can be very different from those of robot designers. Furthermore, technology designs assuming the elderly need to be assisted by machines, rather than use machines to help themselves, reproduce a situation in which the elders' agency is diminished. We suggest a grounded approach to research that would support technology design, accompanied by an iterative practice of collaborative modeling that will include the presentation of research results in the form of ethnographic findings and computational models to elders for reflection and critique.

We would like to increase participation of elderly in design for two reasons:

1. To improve the designs of technology
2. To give elderly more agency in constructing their interactions and environments.

The resulting technology designs would include the viewpoints, needs, and desires of the end users; they would be built according to their understandings of space, interaction, needed and appropriate assistance. Furthermore, we want to build a system that will enable and encourage discussions among the elderly about issues of design, how technologies fit into their communities and everyday lives, as well as about the values and practices that they want to develop in their communities. Finally, we hope that the resulting models and robotic technologies will contribute to the community's ability to self-organize and be reflexive about its change and development.

2. MODELING A COMMUNITY

We imagine the community as an "information ecology" [7]—a social space in which the technology's functions and people's actions and sense-making about robots is mutually constructed. The metaphor of the ecology encourages a focus on the diversity of contexts of use, relevant actors and their roles, and the dynamic changes that a habitat and its denizens go through. Designing for an information ecology also calls for the incorporation of the values and perceptions of community members, as well as the potential differential impacts that the technology may have on various groups. In the case of assistive robots the context of design (i.e. the laboratory) is often different from the context of use in an ecology of care. For example, the design of the assistive robot Paro involves people, spaces, and activities (see Figure 1) that are distinct from those present involved in its use (see Figure 2).

We propose building computational models to observe and think about change in the community. We use agent-based modeling because it allows for us to model interactions and local processes of the community [5]. ABMs such as these can be used not just for design of policies and technologies but also to have a framework for evaluating how the introduction of technologies affects the ecology. Another benefit is that agent-based models produce generative data that gives the ability to not only match real data for validation [6], but also provides easier to understand explanations for why particular patterns emerge across populations [2].

In our models we consider interactions between community members, staff, spaces, and technologies. Interactions can

Figure 1: An office at AIST in Tsukuba, Japan, where Paro, a seal robot for the elderly, is designed.

Figure 2: A nursing home in Japan, the information ecology in which Paro is put to use.

range from conversations to uses of particular technologies. It is possible to generate many kinds of interactions with agent-based modeling software, which is formalized as a type of edge, or link, in a network of interactions. The majority of our selections of technologies, actors, and processes to model will come from interviews and activities with the community. However, we will have to formalize some spaces from top down. In order to do this we consider what objects are most important in supporting community self-organization and self-evaluation. A house, a space that one occupies for much of their time, could be one such object. Common spaces, where members of the community gather and interact, are another relevant space.

The model will be itself designed using grounded methods. We follow carefully what aspects of the ecology are significant to the various actors in it and use those in the model building process. If organizing social events and who participates in events is one of the most important concerns we will focus our model on that. We also document the model building process through field notes to trace the development of design themes and problems throughout the project.

For this poster we will display some prototype models of the retirement home and its community areas and describe the different technical choices we made to code and define each space. We also describe a card game that we are introducing to the members of the community that we will use to gather data. The game is an activity that focuses broadly on the values members ascribe to technology and how these values are negotiated within the group. The games will evolve as the models evolve and, hopefully, help us understand how people makes sense of objects found to be relevant in the models.

3. REFERENCES

[1] R. Birby, F. Broz, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. simmons, K. snipes, A. Schultz, and J. Wang. Designing robots for long-term interaction. In *Proceeding of IEEE International conference on intelligent robots and systems*, pages 2199–2204, 2005.

[2] J. M. Epstein. *Generative social science: Studies in agent-based computational modeling*. Princeton UP, Princeton and Oxford, 2006.

[3] J. Forlizzi, C. DiSalvo, and F. Gemperle. Assistive robotics and an ecology of elders living independently in their homes. *Journal of human computer interaction*, 19:25–59, 2004.

[4] B. Gates. A robot in every home. *Scientific American*, 296(1):58–65, 2007.

[5] N. Gilbert. *Agent-based models*. Sage Publications, Los Angeles, 2008.

[6] V. Grimm and S. F. Railsback. *Individual-based modeling and ecology*. Princeton series in theoretical and computational biology. Princeton University Press, Princeton, 2005.

[7] B. Nardi and V. O'Day. *Information ecologies: Using technology with heart*. M.I.T. Press, 2000.

[8] P. Ruvolo, I. Fasel, and J. Movellan. Auditory mood detection for social and educational robots. In *Proceeding of IEEE International conference on robots and automation*, 2008.

[9] K. Wada and T. Shibata. Social effects of robot therapy in a care house—change of social network of the residents for one year. *Journal of advanced computational intelligence and intelligent informatics*, 13(4):386–392, 2009.

The Jersey Punk Basement Scene: Exploring the Information Underground

Joe Sanchez
Rutgers University
4 Huntington St.
New Brunswick, NJ 08901
00-1-732-932-7500
sanchezj@rutgers.edu

Aaron Trammell
Rutgers University
4 Huntington St.
New Brunswick, NJ 08901
00-1-732-673-3879
aaront@eden.rutgers.edu

Nathan Graham
Rutgers University
4 Huntington St.
New Brunswick, NJ 08901
00-1-910-384-4450
nathang@rutgers.edu

Jessica Lingel
Rutgers University
4 Huntington St.
New Brunswick, NJ 08901
00-1-212-846-8772
jlingel@eden.rutgers.edu

ABSTRACT

This research project attempts to understand the complex information interactions between musicians, promoters, and audience members who perform/promote/transform residential basements into underground music venues. The goals of this project are (1) to obtain an understanding of the information-seeking practices that enable social actors to participate in an underground music scene, (2) to identify and discover the presence or absence of information-communication technologies (ICTs) within the communication network, and (3) to analyze the underground music venues as artifacts situated within a punk-rock subculture.

Keywords

Everyday life information seeking, diversity, deviant information seeking, qualitative methods, punk.

1. INTRODUCTION

Punk culture has traditionally maintained an adversarial relationship with the mainstream. Levine and Steven [1] explains that punk uses mechanisms such as formalized secrecy and indoctrination as a way to preserve itself as a subculture. An example of this formalized secrecy can be found in New Brunswick, New Jersey, where punk-rock shows have been consistently occurring in residential basements, away from mainstream music fans, for over twenty years. Basement parties typically feature known or up-and-coming punk bands, both local and those that drive hundreds of miles, performing in quasilegal venues for little more than donations. Promoters of these shows face both litigation risks from potential accidents and hefty fines for noise-ordinance violations if police are able to locate the shows. Community-driven information needs are at the center of this legally-nebulous, geographically-bound and intensely active intersection of music, subculture and media. Utilizing a transdisciplinary approach, this research project explores the complex information interactions between musicians, party promoters, and audience members who perform/promote/transform residential basements into underground music venues. The goals of this project are (1) to obtain an understanding of the information-seeking practices that enable social actors to participate in an underground music scene, (2) to identify and discover the presence or absence of information-communication technologies (ICTs) within the communication network, and (3) to analyze the underground music venues as artifacts situated within a punk-rock subculture.

2. BACKGROUND

In order to understand punk as a subculture, an appeal must be made to its ideologies, which are frequently analyzed (both from within the community and from without) through the lens of capitalism. Punk appeals to a distinct set of economic goals and aesthetics, and furthermore the punk community, unlike an economic system, must approve the emergent products or texts as punk. Historically, punk labels formed as a way to forgo the "usual commercial division of labor" and vertical market integration [2]. By establishing autonomy through a transformation of exhausted habits and commercialist rituals, punk creates autonomy "through a radical notion of individualism, rather than a (sub)cultural homogeneity" [3]. This push for autonomy, which is reflected in the appropriation and transformation of legitimized symbols by the subculture to resignify the commodified object [4], is a primary agent in the conflict of punk versus capitalism and commodification. The ability to create a new personal narrative and a self-authorized perspective [5] provides those in the punk movement a sense of autonomy. Punk, when theorized as a rejection of capitalism, is continually searching to assign a unique set of use values to objects without a stable system for consumption. As a result of the individualism exhibited, vertical market integration is stifled. The "romanticized isolation" and outright rejection of commodification manifest in punk [6] creates an unbridgeable chasm between the subculture and capitalism.

Also central to the discussion of punk rock is the question of geography [7]. How do localized punk ideals reflect the dissemination of British and American traditions of punk aesthetics and discourse [8]? Discursively, late-seventies British

punk is constructed as postmodern rebellion and read as an attempt to destroy or undermine the status quo of an oppressive capitalist system [9]. Analysis of American, or hardcore, punk is less interested in producing a unified discourse; instead, conversations revolve around American punk as a plural subject requiring the larger framework of its diaspora [10]. When interrogating the commonalities of these literatures, media theorist Michelle Pjilipov [10] asks, how, beyond politics, does music function as a form of cultural exchange? Proceeding from Pjilipov, it is important to interpret the New Brunswick basement scene as a fundamentally discrete phenomenon. It is post-punk (reacting to the mainstream successes of both British and American punk ideologies) and heterogeneous (reflecting the convergence of several things at once). Additionally, it is utopic, dystopic, and geographically bound. These aspects completely differentiate information seeking behavior from topics of hardcore currency, politics, and postmodern ideology. Instead, this is a discussion of identity, resistance, and network.

3. RATIONALE

Within literature that attempts to take a user-oriented approach to human-information behavior, two approaches relevant to this project are to emphasize the social concepts and the situational components at work in information seeking. Both have been developed within the scholarship on everyday life information seeking (ELIS) and will be applied to the information practices of musicians, promoters, and fans participating in the underground music venues. Savolainen's [11] influential paper rejects the categorization of information searchers based on socio-demographic variables, focusing instead on the spatial and temporal circumstances that inform an information-seeking session. These factors are fundamental to McKenzie's [12] work on the information-seeking behavior of women pregnant with twins, which advances a model of information seeking that "reflect[s] the idiosyncrasies present in accounts of ELIS" and identifies four models of gaining information: active seeking, active scanning, non-directed monitoring, and by proxy. Where Savolainen sought to emphasize the context in which individuals acquire information, Chatman [13] has used community as a lens through which to understand information seeking behavior. In order to examine how obtaining information often takes place strictly within the perceived confines of a community, we will investigate the underground music scene in the context of how it is opposed to and outside of mainstream music. Referring to information within a single economic class, Chatman has written that "whether they are seen as seekers of information from others much like themselves or skeptical of claims not personally experienced, the conclusion is that they live in an impoverished information world. This world can be viewed as one that has a limited range of new possibilities, and that other perceptions about reality are not adequate, trustworthy, and reliable" [14]. Knowledge about the basement scene in New Brunswick is bound geographically as well as (sub)culturally, which renders the process of identifying sources for obtaining relevant information less visible and more complex. Fisher, Landry and Naumer's work on information grounds [14] provides a framework to parse this complexity, where their analysis of atmospheres in which information is transferred and obtained includes describing specific environmental factors and social roles that contribute to the utility of gaining knowledge. Ultimately, the authors suggested that "information grounds are fertile territory for non-purposive information behaviour [sic] ... and facilitate the social construction of information needs, and thus, information itself." In identifying the basement music scene in New Brunswick as a community, there is a contingent identification of information needs, which can be (and are) in turn met by a number of different actors operating in different places. As information travels through a cluster of people, the connections made between nodes of data identify gatekeepers of knowledge. In this way, tracing paths of information reveals the stakeholders and power dynamics of a subculture.

4. METHOD

This qualitative study combines interviews, focus groups, and participant observation in order to understand the information seeking practices of participants in the New Brunswick underground music scene. Data collection will be situated within the context of everyday life information seeking behavior [11, 12, 13, 14]. By participating in a natural research setting, the research team can develop a tacit understanding of everyday actions that could be missed or “factored-out” in a lab environment. All members of the research team are musicians and participants in non-mainstream subcultures, and this tacit understanding of the phenomena will help to expose multiple realities within the research setting. By participating in the daily events of a community, participant observation may allow a researcher to see and better understand underlying complexities of these multi-faceted situations. The purpose of observational data in qualitative research is to take the reader into the place that was observed, meaning the reporting of the data will be richly descriptive in-depth and detailed [15].

5. SIGNIFICANCE

Approaching the topic of punk-rock subculture from the perspective of information seeking allows for nuances in the analysis of the ways in which subcultures traffic information, as well as a perspective on information-seeking behavior that analyzes counterculture. This is not to argue that ours is the only method of examining a subculture; it does, however, offer an analytical approach to understanding a topic that is often only considered interpretively. Thus this project aims not only to study information-seeking behavior in a particular subculture, but also to provide a richer understanding of how group identity can be constructed through information.

Tracing patterns of communication that are at once a convergence of subversion, culture, music, and analytical behavior gives us an access point for multidirectional inquiry. When research on information-seeking behavior has examined particular groups, it is often groups belonging to mainstream culture (e.g., office workers, children and young adults in public libraries, or medical professionals). By applying these tools to subculture, we hope to bring a fresh perspective to normative methods and debates. Ideally, this work will function as an intervention in these two discourses, and it should be read in that sense. What fresh perspectives can we bring to two discussions used to traffic normalized ideas?

6. REFERENCES

[1] Levine, H., & Steven, S. (1983) Statements of Fear Through Cultural Symbols: Punk Rock as a Reflective Subculture. Youth and Society, 14(4) 417-425.

[2] Thompson, S. (2001) Market Failure: Punk Economics, Early and Late. *College Literature*, 28(2), 48-64.

[3] Goshert, J. C. (2001) “Punk” after the Pistols: American Music, Economics, and Politics in the 1980s and 1990s. *Popular Music and Society*, 24(1), 87-109.

[4] Dechaine, D. R. (1997) Mapping Subversion: Queercore Music's Playful Discourse of Resistance. *Popular Music and Society* (21)4 p. 7-37.

[5] Traber, D. S. (2001) L.A.'s 'White Minority': Punk and the Contradictions of Self-Marginalization. *Cultural Critique* 48: 30-64.

[6] Temple, Johnny. (1999) “Noise from Underground.” *Nation* 18 Oct: 17-23.

[7] O’Connor, A. (2002) Local Scenes and Dangerous Crossroads: Punk and Theories of Cultural Hybridity. *Popular Music*, 21(2), 225-236.

[8] Lentini, P. (2003) Punk’s Origins: Anglo-American syncretism. *Journal of Intercultural Studies,* 24(2), 153-174.

[9] Davies, J. (1996) The Future of “No Future”: Punk Rock and Postmodern Theory. *The Journal of Popular Culture*, 29(4), 3-25.

[10] Pjillipov, M. (2006) Haunted by the Spirit of ’77: Punk Studies and the Persistence of Politics. *Continuum: Journal of Media & Cultural Studies*, 20(3), 283-393.

[11] Savolainen, R., (1995) Everyday life information seeking: Approaching information seeking in the context of way of life. *Library and Information Science Research* 17, pp. 259 294

[12] McKenzie, P.J. (2002), ``Communication barriers and information-seeking counter strategies in accounts of practitioner-patient encounters’, Library & Information Science Research, Vol. 24 No. 1, pp. 31-47.

[13] Chatman, E. (1991). Life in a small world: Applicability of gratification theory to information-seeking behavior. Journal of the American Society for Information Science and Technology 42(6), p. 438- 449.

[14] Fisher, K., Landry, C., Naumer, C. (2007). Social spaces, casual interactions, meaningful exchanges: "Information ground" characteristics based on the college student experience. Information Research 12(2).

[15] Lave, J., & Wenger, E. (1991). Situated Learning. Legitimate peripheral participation. New York: Cambridge University Press.

The LIS Virtual Library: A case study of library support for an iSchool

Susan E. Searing
Library, University of Illinois, Urbana-Champaign
244 LIS Building
501 East Daniel St, Champaign, IL 61801 USA
001-217-333-4456
searing@illinois.edu

Timothy L. Offenstein
CITES, University of Illinois, Urbana-Champaign
1120 Digital Computer Laboratory
1304 W. Springfield AvenueUrbana, IL 61801 USA
001-217-244-2700
timo@illinois.edu

ABSTRACT

What impact has the iSchool movement had on the collections and service programs of the libraries at universities that are homes to iSchools? How are academic libraries meeting the information needs of iSchool faculty and students? At the University of Illinois, Urbana-Champaign (UI), the expansion of the curriculum and research agenda of the Graduate School of Library and Information Science (GSLIS) to encompass the far-reaching perspectives of an interdisciplinary iSchool influenced the development of a new service model for library support. The library was also challenged to support GSLIS's very successful online MLIS and Certificate of Advanced Study degree programs.

At the UI, a system of distributed, departmental libraries has been in place since the 19th century. A separate Library & Information Science (LIS) Library was housed in the Main Library facility from the 1920s until May 2009, when its collections were merged into other libraries. The new model for LIS library services combines a more robust virtual presence with an intensified human presence in the GSLIS building. These changes are part of a much larger initiative to create a more flexible organizational structure for the University Library overall – a structure that recognizes the increasingly interdisciplinary nature of academic inquiry, the critical importance of digital information resources, and the opportunities for collaborative approaches to the provision of library services and collections using information technology. [1] Over the past decade, these themes have been echoed repeatedly in studies of library use, scholarly information-seeking, and the future of the academic library. [2] Recent writings have also affirmed the value of subject specialist librarians and library services targeted to communities of scholarship and practice. [3]

This case study of a change process that is still in progress highlights the tensions and opportunities that are created for the library system when an academic unit shifts and enlarges it scholarly focus. At the UI, evidence (both quantitative and qualitative) about library user behavior and needs was brought to bear on the decision-making and planning processes. Members of the UI's iSchool community were involved, as well as members of the University Library's faculty and staff, in charting a future for library collections and services for LIS and related fields. For example, the School's Associate Dean served on the planning team, and cataloging instructors advised on the retention and relocation of reference sources for teaching and performing cataloging.

Most notably, a team consisting of the LIS Librarian and two students, within the traditional framework of an independent study course, spearheaded the transformation of the LIS Library's former web site into a "virtual library" and portal to disciplinary information. [4] The enrolled students had individual learning goals which were met by readings, consultations with experts, and hands-on design and problem-solving work. The team conducted a round of usabilty testing, using Morae software, which led to significant design revisions before the site went live. Since the launch of the website in August 2009, further valuable critiques have been offered by students in the online GSLIS course, Interfaces to Information Systems.

While the site design is constrained by the limitations of the Library's content management system (OpenCMS) and locally mandated design templates, the University Library places almost no restrictions on content or organization of subject-specific websites within its domain. Therefore, the design team was able to organize the information in ways helpful to the UI iSchool community and the wider audience of information professionals on campus. The CMS structure permits easy updating and modification of the site. The site itself provides opportunities for users to suggest additional content and/or comment on the design. The site incorporates newly-acquired resources and tools, such as the "LIS Easy Search," a federated search across three major LIS journal indexes, local and consortial online catalogs, ebook sources, and the UI's institutional repository. This new feature was developed after the planning team undertook a survey of LIS Library users, which revealed that the virtual service most valued by students is access to LIS-specific databases. In contrast, the same survey showed that faculty users of the LIS Library online valued most highly its virtual new books shelf. Because new LIS books were no longer directed to a single physical location after the closure of the LIS Library, the

new virtual library incorporates a complex search of the University Library's database of new books, in order to gather together records for LIS titles. Other services of the LIS Virtual Library include a news section and an expanded, subject-categorized set of links to e-resources, both licensed and open access. Some new content aims to bridge the gap between the user's experience of the old physical library and the new virtual one--for example, pages explaining where LIS printed books and journals may now be found in the UI's vast library system.

Like most academic libraries, the LIS Library at UI has been shifting from mostly print to mostly digital collections over the past several years. On the one hand, the closing of the physical library was simply an inevitable evolutionary stage, responsive to increasing digital publishing and changing modes of scholarly information-seeking. On the other hand, many users experienced the closure as a decisive and even tectonic shift in their personal information worlds. Relevant print materials are still collected and housed in appropriate campus libraries, but the LIS Virtual Library is now, for most purposes, _the_ library for the iSchool. The increased presence of the LIS Librarian in the GSLIS building helps to keep users connected with the resources of the University Library and bridges the gap between the old service model and the new.

The transformation of the LIS Library demonstrates how a successful transition from a traditional information service model to a new one must be grounded in knowledge of the unique needs and customs of the library, university, and population of users. Further, it demonstrated that the University Library can involve iSchool students in meaningful projects that both further their learning and contribute to the improvement of information services. The evolution of library services is a noticeable impact of the iSchool movement, as the UI example proves.

Categories and Subject Descriptors

H.3.5 [**On-line information services**]: Web-based services.

H.3.7 [**Digital libraries**].

K.3.2 [**Computer and information science education**]

General Terms

Management

Keywords

User-centered design; digital library; federated search; usability; information behavior; content management system; academic libraries; embedded librarians; evidence-based librarianship; iSchools; library and information science.

REFERENCES

[1] "New Service Model Programs," URL= http://www.library.illinois.edu/nsm/

[2] See, for example: De Rosa, C., Cantrell, J., Cellentani, D., Hawk, J., Jenkins, L. and Wilson, A. 2005. Perceptions of Libraries and Information Resources: A Report to the OCLC Membership. URL=http://www.oclc.org/reports/2005perceptions.htm; Housewright, R., & Schonfeld, R. 2008. Ithaka's 2006 studies of key stakeholders in the digital transformation in higher education. New York: Ithaka. URL=http://www.ithaka.org/research/faculty-and-librarian-surveys; Palmer, C. L. , Teffeau, L. C., & Pirmann, C. M. 2009. Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development. OCLC Programs and Research. URL=http://www.oclc.org/research/publications/library/2009/2009-02.pdf; McNicol, Sarah. 2003. LIS: the interdisciplinary research landscape. Journal of Librarianship and Information Science 35(1), 23-30; Tenopir, C. 2003. Use and users of electronic library resources: An overview and analysis of recent research studies. Washington, DC: Council on Library and Information Resources. URL=http://www.clir.org/pubs/abstract/pub120abst.html.

[3] Special issue on liaison librarian roles. 2009. Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC , no. 265, 1-33; Cassner, M., & Adams, K.E. 2008. The subject specialist librarian's role in providing distance learning services. Journal of Library Administration 48(3/4), 391-410; Bartnik, L. 2007. The embedded academic librarian: the subject specialist moves into the discipline college. Kentucky Libraries 71(3), 4-9.

[4] "Library and Information Science (LIS) Virtual Library." URL=http://www.library.illinois.edu/lsx

Save the tweets so you can understand the birds

[Extended Abstract]

Claudia Serbanuta
University of Illinois
501 E Daniel Street
Champaign, Illinois
cserban2@illinois.edu

Tiffany Chao
University of Illinois
501 E Daniel Street
Champaign, Illinois
tchao@illinois.edu

Aiko Takazawa
University of Illinois
501 E Daniel Street
Champaign, Illinois
aikot@illinois.edu

ABSTRACT

The emergence of Web 2.0 technologies in the digital age offers increased opportunities for worldwide information dissemination and self-expression. However, the capture, archive, and future use of this user generated content as a scholarly resource has been seldom addressed in the either the library or archive context. This poster presents a case study regarding the virtual exchanges and events that followed the April 2009 Parliamentary elections in Moldova. The civil unrest in the ex-Soviet republic was first presented in the international media as a "Twitter revolution" and the buzz 'web 2.0' label stayed with the events as they garnered the increased attention of a global public (see references).

The events, the information flow and the emotional outpouring surrounding the Moldova election were primarily manifested through Twitter. The community-driven use of this online platform to convey instantaneous coverage illustrates the increasing use of digital spaces as a public extension for social interaction. Introduced in 2006, Twitter has become one of the most popular internet applications for personally conveying brief textual messages to a group of potentially interested individuals (microblog); preliminary research has been conducted regarding the intentions for using this type of microblog by individuals and the communities formed (Java, 2007). The aggregation of these messages (tweets) based on a common identifier (hashtag), which links with an issue or event, presents an invaluable record of social memory with immense contextual and research value. Tweets containing the hashtag created for Moldova events (#pman) were archived by the authors. Using this raw data, we will explore the value of such information from an interdisciplinary, scholarly perspective. By analyzing the #pman archive, this poster will address the following questions: What does the content in the archived messages tell us about the events from Moldova and the people involved? Who might use such an archive of user-contributed content in the future? How is preservation of this type of digital content made possible?

iConference 2010, Champaign, Illinois

Through the investigation of the archived Twitter content, we will explore what these messages tell the public about the events in Moldova. We will look at the depth of data contained in tweets and in the meta-information available through Twitter. The content will be examined for information on how communication and information sharing transpired. Special attention will be given to: the language used (Romanian, English, or Russian) in the same online space to describe the events; the structure of the messages (how the statements were constructed and how they circulated) and hyperlinks (the outside information sources that were attached to the posted messages, i.e. URLs to multimedia sources). Patterns, observations and further areas of inquiry will be presented in the attempt to establish the value of preserving such a unique data set.

Based on its social and contextual richness, this data set would be of interest to scholars from a variety of fields. During the election and subsequent protests, this content became part of not only the local but also the global community memory. It also serves as an essential resource which documents the entire scope of events for future users to confront and interpret. Our analysis of the Moldova election response demonstrates a potential application for the development of a civic society by recognizing community memory and identity, and its lasting impact on social participation and democracy. From this perspective, the individual messages and images shared have sustainable value in the aftermath of such intense social and historical events. Inherited values of the community as well as memories of the event remain alive and continue to emerge at different levels of society as well as in individual thought, patterns which could be potentially traced through the preserved record of the #pman archive.

With the limitation of 140 characters, tweets and other types of user-generated content in Web 2.0 environments have generally been considered ephemeral in nature where long term preservation of such information has not been observed. Our case study illuminates an opportunity to learn more about the preservation needs for such unique content by examining the process and the types of information that have been captured (i.e. time stamp for when tweet was posted, native language of the tweet) while also considering the additional details and components that would make this data set more robust for potential reuse by another researcher. The archiving initiatives of electronic mail that have been undertaken by government, institutional organizations, and repositories along with the emergence of online archival services that specialize in Twitter-produced content

(i.e. www.twapperkeeper.com) will serve as a comparative framework for determining the preservation criteria of our dataset (Lukesh, 1999).

As part of the microblogging and Web 2.0 phenomena, Twitter presents a crucial reflection of modern social culture in the digital realm. Our exploratory case study of the Moldova "revolution" shows the importance of this online venue in its capability to document informal exchanges from a worldwide network of participants on a variety of social, political and cultural issues and events as they occur and unfold. The content generated through Twitter helps to strengthen the knowledge of specific historical and social contexts if attention is directed to its preservation. Investigation of the #pman archive will reveal the depth of community participation and the impact of such a resource for long-term study while introducing a new primary source for preservation consideration.

Following the Moldova elections in April, another "Twitter Revolution" complemented the Iran election protests in June 2009. The international media and public reactions were even more prominent and pervasive- this time, tweets were more efficient in carrying these messages around the world as seen in the increased number of internet users archiving tweets with #iran and #iranelections. Through the construction of these archives and the provision of public access to them, the inherent value of this content is not only reflective of those thoughts and emotions expressed by the people behind the tweets but also of those individuals who sought to protect the ideas they embody.

Categories and Subject Descriptors

H.4.3 [**Information Systems Applications**]: Communications Applications; H.3.5 [**Information Storage and Retrieval**]: On-line Information Services—*Web-based services*

General Terms

Documentation, Reliability

Keywords

Twitter, digital archives, social networking

1. REFERENCES

[1] A. Applebaum. The twitter revolution that wasn't. `http://www.washingtonpost.com/wp-dyn/content/article/2009/04/20/AR20090%42002817.html` Accessed on 20 April 2009.

[2] R. F. Europe. Moldova's "twitter revolution". `http://www.rferl.org/content/Moldovas_Twitter_Revolution/1605005.html` Accessed on 8 April 2009.

[3] N. Hodge. Inside moldova's twitter revolution. `http://www.wired.com/dangerroom/2009/04/inside-moldovas/` Accessed on 8 April 2009.

[4] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. 2007.

[5] S. Lukesh. E-mail and potential loss to future archives and scholarship or the dog that didn't bark. *FirstMonday*, 4(9), 1999.

[6] E. Morozov. Moldova's twitter revolution. `http://neteffect.foreignpolicy.com/posts/2009/04/07/moldovas_twitter_re%volution` Accessed on 7 April 2009.

The Value of Public Sector Information as a Strategic Resource for Socioeconomic Development Research and Policy Activities in South Africa

Raed M. Sharif
Ph.D. Candidate in Information Science and Technology at the iSchool of Syracuse
3 Lowry Sq. Scarborough, ON. M1B1N7. Canada.
+1-647-567-2212
rmalshar@syr.edu

ABSTRACT

Although it has always been an important asset to those who posses it, in the current knowledge society, information is considered as one of the most important goods in our daily life (Machlup, 1962; Porat, 1977; Mueller 1995; Stiglitz, 2000).

At the same time, the public sector is the biggest single producer and owner of a large variety of information (e.g., health and geographic information, financial reports, social and economic statistics, legislation and judicial proceedings, food and water resources information, and many other kinds of data and information, collectively referred to as Public Sector Information). Public Sector Information (PSI) represents an important resource with vast socio-economic potential to different communities.

According to Horton (2002) diffusing public information and knowledge resources efficiently and effectively is essential to:

- "Sustaining the competitive competency of the country's businesses and industries, in both domestic and global marketplaces;
- Attaining the highest levels of educational excellence for all the nation's children and adults in a lifelong learning context;
- Enabling citizens to participate more effectively in all facets of a democratic society;
- Informing public officials at all levels of government so that they can enact better laws, formulate and enact enlightened public policies, monitor the programs they authorize effectively, and govern fairly, equitably, and wisely; and,
- Enhancing the quality of life of all a country's citizens, including responsibility to the special government information needs of disadvantaged and disabled individuals." (p.3)

For example, governments can use this strategic resource to make sound policies and to promote transparency and accountability; and private sector can use it to produce innovative products and services, which in turn can contribute to the nation's economy. As for citizens, PSI is essential for exerting their civic rights and enabling democratic participation. For civil society organizations, PSI can be a strategic resource for their work, especially in areas such as poverty eradication, public health, food security, disaster management, and governance, where the combination of different types of PSI (e.g., geo-spatial, economic, and health data) can be of tremendous value for successful targeting and support of marginalized communities.

Finally, scientific and policy research communities benefit tremendously from the PSI. The list of benefits to the community includes the promotion of interdisciplinary, inter-sector, inter-institutional, and international research. It is especially important for these communities when their research and activities are focused on socioeconomic development issues. My study focuses on this last community.

In this poster presentation I share and discuss the preliminary analysis and initial findings from the data I collected during my six months fieldwork in South Africa for my dissertation titled: The Value of Public Sector Information as a Strategic Resource for Socioeconomic Development Research and Policy Activities in South Africa.

In this study I investigate whether, and if so how, PSI is utilized by South African organization working in the area of socioeconomic development research. More specifically, my study aims to answer the following questions:

- To what extent and in what ways is PSI utilized by research organizations?
- What characteristics and conditions of the PSI facilitate or hinder its acquisition and assimilation?
- What organizational conditions enable successful exploitation of PSI by these organizations?

Employing a qualitiative, multiple-case approach (Yin, 2003), I draw upon literature from the fields of economics of information and organizational studies. I use literature from economics of information to understand the differences between public vs. private information and the importance of external information to organizational innovation; and literature from organizational studies, particularly about absorptive capacity (Cohen & Levinthal, 1990), to understand the organizations' processes to identify, acquire (including factors that facilitate or hinder access and acquisition), assimilate, and exploit this strategic resource. My case studies explore how these organizations transform PSI

from a source of potential value to a source of actual value to their socioeconomic development research.

I also use ideas from organizational learning (Argyris & Schön, 1978; Huber, 1991) and organizational innovation (March & Simons, 1958) literatures to help me better understand the organizational conditions (internal and external) for successful utilization of the PSI. Mainly qualitative data were collected from these research organizations through in-depth, semi-structured interviews and document analysis over a period of six months.

My initial analysis suggests that the PSI represents a very strategic resource to the development work of these organizations (especially for what purposes and in what areas this resource is being used), uncovers very important obstacles facing these organizations in identifying, acquiring, and utilizing the South Africa PSI, and documents some best practices and organizational factors that facilitate these processes. It is expected that the discussions and findings of this study will have theoretical and policy contributions, and will be of special importance to organizations working in the area of socioeconomic development research, the government of South Africa (and hopefully governments in other developing countries), and subsequently to the people of South Africa.

Categories and Subject Descriptors

eGovernment: information policy, economics, ethics, law, technologies of privacy and trust

Keywords

Public Sector Information, Socioeconomic development, Absorptive Capacity, South Africa, Developing Countries

REFERENCES

[1] Argyris, C., & Schon, D. (1978). Organizational Learning: A Theory of Action Perspective. Reading, MA: Addison-Wesley.

[2] Cohen,W.M., and Levinthal,D.A., (1990). Absorptive Capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1) 128-152

[3] Horton, F.W., (2002). Public Access to Government Information and Information Literacy Training as Basic Human Rights, White Paper prepared for UNESCO, the U.S. National Commission on Libraries and Information Science, and the National Forum on Information Literacy, for use at the Information Literacy Meeting of Experts, Prague, The Czech Republic.

[4] Huber, G.P. (1991). Organizational learning: The contributing processes and the literatures. Organization Science, 2(1) p. 88-115.

[5] Machlup, F. (1962). The Production and Distribution of Knowledge in the United States, Princeton UP.

[6] Mueller, M. (1995). Why Communication Policy Is Passing " Mass Communication" By: Political Economy as the Missing Link. Critical Studies in Mass Communication, 12 (4) 457-72.

[7] March, J. G. and Simon, H. A. (1958). Organizations. John Wiley & Sons.

[8] Porat, M. (1977). The Information Economy: Definition and Measurement. Office of Telecommunications, Washington DC,

[9] Stiglitz, J. E. (2000). The Contributions of the Economics of Information to Twentieth Century Economics, Quarterly Journal of Economics, 115(4), pp. 1441-1478.

[10] Yin, R. (2003). Case Study Research, 3rd ed. Sage Publications, Thousand Oaks, CA.

Why do users neglect suggestions?: Effects of semantic relatedness and task on word recognition

C. L. Smith & N. Wacholder

c.smith; ninwac @ rutgers.edu

We report work in progress on the question "Why do searchers frequently fail to use potentially valuable query suggestions?" [1,2]. We hypothesize that failure is due, at least in part, to interference with the searcher's ability to recognize a semantic relationship between the words used in a query and the words in a suggestion. In our study, we measure *semantic priming* as an indicator of a searcher's recognition of relationships between words. This poster presents preliminary results from one experiment in the study.

1. INTRODUCTION

Generally, our research objective is to investigate people's recognition of related words in the context of interaction with a search system. More specifically, we are interested in how the tasks of formulating a query or scanning a results page affect recognition. In our broader study, we approach these questions in a series of controlled experiments that isolate effects due to factors such as semantic relatedness, context, and task.

This abstract and the poster are organized as follows. First, we briefly define and describe the principal element of our methodological approach: *semantic priming*. Next, we describe our baseline study, which uses a standard approach for measuring semantic priming, the *lexical decision task*. Then, we describe a new experimental approach, which is designed to invoke a decision task that occurs in the course of interactive search. Our poster presents results from an experiment conducted using this new task.

2. SEMANTIC PRIMING

Semantic priming is a well-established, extensively investigated cognitive phenomenon [3]. Psychologists and linguists use measures of semantic priming in a wide range of studies, including areas such as memory, reading, and perception. Semantic priming refers to an increase in the availability of a word in memory, where the increase is caused by the processing of a preceding, semantically related word or other stimuli such as an image. For example, the word *kitten* "primes" the semantically related word *cat*; the unrelated word *table* does not prime *cat*. The difference in availability is termed the *semantic priming effect.* There is a large literature on the many factors that affect semantic priming. In our experiments, we manipulate semantic relationships between words, the order of words, and the subject's task, as independent variables. We measure *semantic priming* as the dependent variable.

3. BASELINE STUDY

In our baseline experiment, we used a standard methodology for measuring semantic priming: the *lexical decision task* (LDT). During one iteration of this task, a volunteer sees a sequence of computer screens (see Figure 1). The first screen displays a fixation point, which draws the volunteer's eye to the center of the screen. Next, a real English word is displayed very briefly (~150 milliseconds); because it is processed first, this word is called the *prime*. A blank screen then flashes very quickly (~50ms). Finally, a second string of letters is displayed; this string is called the *target*. The target can be a real English word or a pronounceable non-word. The volunteer must decide very quickly (within 1 second) whether the target is a real English word (the *lexical decision*). The volunteer indicates the decision by pressing one of two buttons. The time taken to press a button is called the *response time* (RT).

For each iteration of the task, a volunteer may experience one of three possible target conditions:

- **Related-word**: the target is a real word, and the prime is related to the target
- **Unrelated-word**: the target is a real word, and the prime is unrelated to the target
- **Unrelated-nonword**: the target is a nonword

Our baseline measure of semantic priming compares response times under the related-word and unrelated-word conditions. The semantic priming effect is the difference between mean response times under the two conditions.

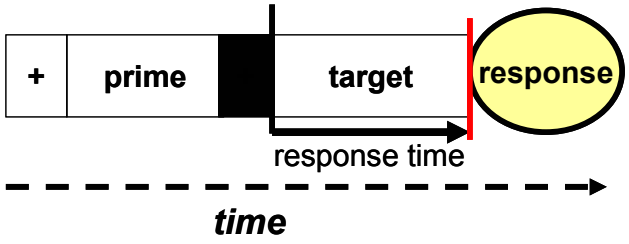


Figure 1. Sequence of screen displays and response in lexical decision task

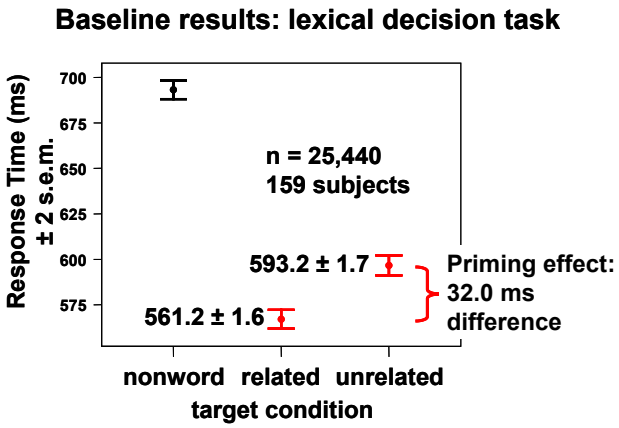


Figure 2. Semantic priming effect in baseline experiment

159 volunteers participated in our baseline experiment, each of whom completed 160 lexical decisions. We find a statistically significant difference in response time for target conditions ($F(2, 25280)=2087$, $p<.001$). Post-hoc analysis using Scheffe's test indicates that response time is significantly different in each of the three target conditions. On average, response time was 32 milliseconds faster in the related-word condition (see Figure 2, above).

4. THE EFFECT OF TASK ON PRIMING

In the design of our larger study, we assume that the words in a searcher's query "primes" the recognition of words in a subsequent display. In this way, we relate primes to query-terms and targets to words displayed in query suggestions. The results reported in our poster show how semantic priming is affected by the task of looking for a word on a two-item list.

For this work, we have used target displays that contain two strings, with one displayed above the other on the screen. For each iteration of a task, a volunteer experiences one of 14 possible target conditions. Table 1 shows an example of the possible target conditions for the prime word *cat*. We have developed a new experimental task, which we call a *presence decision task* (PDT). The task is similar to the LDT, except that rather than deciding whether the target contains a real word, volunteers decide whether the prime word is *present* in the target. Volunteers participating in the reported experiment were assigned randomly to either the PDT (n=55) or the LDT (n=51). For the LDT, volunteers decided whether *both* strings in the target display were real English words. We know of no other semantic priming study that has examined the effect of searching for the prime word *within* the target display.

Table 1. Example of target conditions for the prime word "*cat*"

position in target display	BOTTOM STRING IN TARGET				
		repeated prime	*related word*	*unrelated word*	*non-word*
TOP STRING IN TARGET	*repeated prime*	---	cat kitten	cat army	cat glive
	related word	kitten cat	kitten tiger	kitten army	kitten glive
	unrelated word	army cat	army kitten	army table	army glive
	non-word	glive cat	glive kitten	glive army	---

This work is funded by a generous grant from Google.

5. REFERENCES

[1] Kelly, D., & Fu, X. (2006). Elicitation of term relevance feedback: An investigation of term source and context. In Proceedings of the 29th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '06), Seattle, WA, 453-460.

[2] Kelly, D., Gyllstrom, K., and Bailey, E. W. (2009). A comparison of query and term suggestion features for interactive searching. In *Proceedings of the 32nd international ACM SIGIR Conference on Research and Development in information Retrieval* (SIGIR '09), Boston, MA, 371-378.

[3] McNamara, T. (2005). *Semantic priming: Perspectives from memory and word recognition.* New York, New York: Psychology Press.

Applying multimodal discourse analysis to study image-enabled communication

Jaime Snyder
School of Information Studies,
Syracuse University
245 Hinds Hall
Syracuse, NY 13203
jasnyd01@syr.edu

ABSTRACT

A multimodal analytic framework is introduced to contribute to a discourse-oriented study of the creation of visual information. While much visually based research focuses on the image artifact, an ongoing study seeks to shed light on the phenomenon of image creation as a communication practice. This requires a content analytic methodology capable of addressing issues related to modalities of expression and their interaction, co-occurrence and co-deployment during exchange of meaning. Multimodal discourse analysis, an emerging area of discourse studies, is proposed as a valuable contribution to the current study and to the field of information science in general.

1. INTRODUCTION

Anyone who has reached for pen and paper during a discussion in order to clarify a thought or prompt a response from a companion has exploited the potential of image-making to enhance communication. Because images and visual information enable exchange of meaning across a range of contexts, they play an increasingly important role in how we work and communicate, in both face-to-face and virtual environments.

Marks on a napkin or sketches on a white board are information artifacts embodying a particular type of image-enabled communicative practice that plays a specific role in the exchange of meaning between individuals. These spontaneously created visualizations can anchor, bridge, and facilitate the flow of information at crucial moments in a conversation. *Image-enabled discourse* is the term used to refer to this phenomenon in a broad sense, and *ad hoc visualizations* (i.e. napkin drawings) are one type of image-enabled practice.

In the context of conversations, image creation is an interactive process that generally results in the creation of annartifact, but also includes the motivation or need for the image, the deployment of the image in a specific context, and the reception of the image within an overarching communicative structure [1]. When we focus predominately on the content of that image (either through automated analysis or more qualitative interpretation), we run the risk of generating static analyses of graphical content in which the image is seen as a *fait accompli*, rather than evidence of an interactive process of communication. By improving our understanding of the dynamics of image-enabled conversations, we can build better tools to help people engage in more effective communication

The purpose of this poster is to introduce two approaches to the analysis of multimodal communication that can provide the necessary framework for an investigation into the creation of visual information during face-to-face conversation. Norris' approach to multimodal interactional analysis and Baldry and Thibault's techniques for multimodal text transcription both provide guidance for the operationalization of analytic methods necessary for a discourse-oriented study of image-enabled communication.

2. RESEARCH QUESTIONS

Recent research focused on the creation of images within the context of face-to-face conversation seeks to address the following research questions:

- RQ1: In what ways does the ad hoc creation of images contribute to communication during face-to-face conversations?
- RQ2: What communication strategies are associated with the creation and use of ad hoc visualizations in these interactions?
- RQ3: Which characteristics of visual modes of communication are most salient to the use of ad hoc visualization strategies?

Addressing these questions requires a perspective that recognizes the constructed and dynamic nature of the exchange of meaning between individuals across a variety of modalities. Multimodal discourse analysis, an emerging methodology in the field of discourse linguistics, is introduced to provide an analytic basis for this investigation.

3. BACKGROUND

Although no branch of discourse studies specifically invites extending theories into the realm of image-enabled communication, an emerging subfield of discourse linguistics is highly relevant for this investigation. Multimodal discourse analysis is the study of the intersection and interdependence of various modalities of communication within a given context. Researchers in this area seek to identify the influence of mode on

meaning within a given context, focusing on co-occurrence and interaction between multiple semiotic systems [2].

Generally speaking, *mode* refers to a distinct semiotic system for expressing meaning using specific conventions [3,4]. At the heart of most work in the area of multimodal discourse is the principle that communication occurs across more than a single mode and is therefore inherently *multimodal* [4.5].

According to social semiotic theories related to modality, the form or format of an expression plays a specific role in the communicative power of a sign [4.5]. Modality cues are used in the creation of meaning, referred to as "motivated signs" in the vocabulary of social semiotics [5]. And "…any semiotic mode, even smell, can be conceived of as a loose collection of individual signs, a kind of lexicon, or a stratified system of rules that allow a limited number of elements to generate an infinite number of messages" [6] although the precise nature of those rules or conventions of use can be more or less specific [4].

The generalized definition of modality as a unique semiotic system allows us to go beyond superficial distinctions (such as media or format) to a more complex understanding of how a mode of expression affects the contextualized exchange of meaning(s). Norris highlights this, stating "different communicative modes possess different materiality" [7]. She attributes audible materiality to spoken language, despite its being neither visible nor enduring, while gesture is visible, but also quite fleeting. Print is visible and enduring, as is physical layout of objects. In the context of image-enabled discourse, there are properties of ad hoc visual communication, specifically mark-making, that are uniquely suited to providing the requisite conditions for a person to employ certain communicative practices. For example, drawing naturally has the affordances of being persistent, tangible and visible.

4. ANALYTIC FRAMEWORKS

Two specific approaches to analyzing multimodal interactions can be applied in an analysis of image-enabled conversations to identify salient aspects of communication and the role that modality plays in the exchange of meaning.

4.1 Multimodal interactional analysis

In her approach to the analysis of the interaction of multiple modes of communication is a single context, Norris focuses on "what individuals express and react to in specific situations, in which the ongoing interaction is always co-constructed" [7] She is not just looking at verbal expressions, but at other types of indicators such as head position, body position and layout of objects or spaces to reveal the ways in which this co-constructed is built. Her analysis of multimodal interactions is based on the concepts of *awareness* and *attention*. She clearly states that "Awareness/attention comes in degrees, and a person may be phenomenally aware of something without paying much attention to it."

One of the key concepts offered by Norris that can help clarify the unique communicative qualities of ad hoc visualizations and begin to help us understand how this mode interacts with other types of expressions is the distinction between *embodied* and *disembodied* modes of communication. Music, for example, can be either embodied or disembodied. If a radio is playing in the background while a couple is sitting at a table having breakfast, music is seen as a disembodied mode of communication, not receiving a lot of attention, but the couple will probably have some awareness of it. However, if one of the participants in the conversation breaks into song, music becomes an embodied form of communication, bringing a different level and degree of attention and awareness, not unrelated to the fact that this could be seen as an unusual or unconventional occurrence.

4.2 Multimodal transcription

Baldry and Thibault's approach to multimodal text transcription follows Halliday's definition of text as "living language" regardless of whether it is spoken, written or takes another medium of expression [8]. They recognize that "Different semiotic modalities make different meanings in different ways according to the different media of expression they use."

Multimodal transcription is specifically designed to retain relationships between unique modes of expression in order to retain evidence of differences as well as things like co-deployment. It allows *phasal* expressions (i.e. time- or series-based expressions such as gesture) to be recorded and transcribed alongside *clustered* expressions (i.e. groupings or sets of static expressions such images in a magazine spread or elements on a web page).

Additionally, Baldry and Thibault's techniques highlight the multi-level aspects of meaning making across modalities. They use the concepts of *context of situation* and *context of culture* to delineate important distinctions between modalities. They also examine relationships between individual multimodal *texts* and multimodal *genres*. And like Norris, they acknowledge that *primary* and *secondary* genres exist within any multimodal text.

5. CONCLUSION

Within the field of information science, there is need for more robust methodology that addresses the role that information creation (and visual information, in particular) plays in the exchange of meaning during interactions between individuals. There is great value in having the ability to differentiate images not just by form or format but also by role in communication.

The methodological approaches presented here, multimodal interaction analysis and techniques for multimodal transcription, provide a practical basis for investigating the role drawing plays in conversation. Originating from a discourse perspective, these approaches capture the interaction and co-deployment of multiple modes of expression, and allow researchers to track the influence of these differences in exchange of meaning. Use of these methodologies will help to increase understanding of image-enabled communication and allow us to better exploit this innate human communication practice when building tools and systems.

6. REFERENCES

[1] Ware, C., Information Visualization: Perception for Design. 2000, San Francisco: Morgan Kauffman.

[2] Royce, T.D. and W.L. Bowcher, eds. New Directions in the Analysis of Multimodal Discourse. 2007, Lawrence Erlbaum Associates, Publishers: Mahwah, NJ.

[3] Bateman, J.A., Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents. 2008, New York, NY: Palgrave Macmillan.

[4] Kress, G. and T.V. Leeuwen, Multimodal Discourse: The Modes and Media of Contemporary Communication. 2001: Arnold.

[5] Kress, G. and T.V. Leeuwen, Reading Images: The Grammar of Visual Design. 2006: Routledge.

[6] Levine, P. and R. Scollon, Discourse and Technology: Multimodal Discourse Analysis. 2004, Washington, D.C.: Georgetown University Press.

[7] Norris, S., Analyzing multimodal interaction: A methodological framework. 2004, New York and London: Routledge.

[8] Baldry, A. and P.J. Thibault, Multimodal transcription and text analysis. 2006: Equinox.

Common Ground: Exploring the intersection between information, technology, art and design

Jaime Snyder*
Michael A. D'Eredita
Robert Heckman
Jeffrey M. Stanton
School of Information Studies,
Syracuse University
245 Hinds Hall
Syracuse , NY 13203

* jasnyd01@syr.edu

ABSTRACT

University research is becoming increasingly multidisciplinary in both the nature of the problems being investigated and the make-up of the teams of researchers that tackle these complex challenges. Information schools are in a unique position to participate across a range of these projects. This poster describes an initiative to discover potential areas for collaboration between Syracuse University's iSchool and the College of Visual and Performing Arts, focusing on the synergies between information, technology, art and design.

1. INTRODUCTION

Multidisciplinary research can take many forms. Recognizing a strong synergy between information, technology, art and design, the iSchool at Syracuse Universoty has engaged in an informal partnership with the university's College of Visual and Performing Arts (VPA). The two schools are spearheading a year of exploration entitled *Common Ground*, referring to the rich and fertile territory encompassed by multidisciplinary projects, especially those that bring together information technology and the creative arts.

The goals of the initiative include:

1) Engaging and enlightening faculty and students about the benefits of multidisciplinary approaches to creative problem solving.

2) Providing a foundation for future programs such as dual majors, certificate programs and new multidisciplinary majors.

These goals are being reached through a series of special projects, presentations and classroom activities that will bring together researchers, designers and artists to focus on tough real world problems that could be solved through multidisciplinary thinking.

2. SELECTED PROJECTS

2.1 Innovation Studio

The Innovation Studio was created to facilitate teaching and learning inspired by the creative arts fields, where the studio is the hub for creative experimentation, collaboration and community building. The goal of the Innovation Studio is to enable iSchool faculty and students to engage in hands-on, design-focused learning experiences.

Studio learning generally involves one or more of the following:

- Posing problems to students that might not have just one "correct answer"
- Allowing students the freedom to explore possible solutions independently, according to their own paths of development
- Encouraging regular exchange of constructive criticism between peers
- Maximizing the time spent working in a communal setting where successes and failures are shared with all the participants in the class
- Allocating a minimal amount of time to lectures in exchange for more time spent designing, evaluating, improving and fabricating creative products

The iSchool recognizes that fostering this type of environment includes dedicating a space equipped with appropriate furniture, hardware and software, and also developing a culture that explicitly values creativity and innovation. As part of the *Common Ground* initiative, faculty from art and design are engaging with iSchool instructors to share their experiences and expertise in studio-based practices.

2.2 The iSchool Windows Project

Launched in April 2009, the iSchool Windows Project began as an open call for proposals for the creation of six site-specific artworks to be installed in window wells located in the ground floor of the building that houses the School of Information Studies. The initial success of this unique collaboration with colleagues in the art and design fields paved the way for expansion of the *Common Ground* initiative.

The project's commission sought to raise awareness of the iSchool's accomplishments and scholarship; explore connections

between information, technology, art, and design; provide a commission opportunity for university artists; and build on the expanding partnership between the iSchool and the College of Visual and Performing Arts.

Due to an overwhelming response from the university's artistic community, a total of eight commissions were awarded for original art works to be permanently installed throughout the building. The artworks feature a range of materials, including ceramics, video, fiber arts, metal, and acrylic plexiglass. The pieces have become part of the university's permanent collection and are showcased in six ground floor window wells, on video monitors throughout the building, and in the first-floor lobby area.

2.3 Multidisciplinary courses

The iSchool and VPA have jointly hosted several multidisciplinary course offerings over recent semesters. A selection of these courses is described below. Both schools hope that the *Common Ground* initiative will yield more opportunities for instructors and students to collaborate together in the classroom.

2.3.1 What's the Big Idea and Idea2Startup

A two-part course, spanning fall and spring semesters, provides students with the skills and support to develop their own entrepreneurial ventures, with topics of instruction ranging from jump-starting the creative process to patent searches. Designed by iSchool faculty member Michael A. D'Eredita, *What's the Big Idea?* and *Idea2Startup* are co-taught by instructors from the iSchool, VPA's Department of Design, the School of Management, College of Engineering and entrepreneurs from industry. The first part of the multidisciplinary course, entitled *What's the Big Idea?,* focuses on developing and building student ideas, informed by market forces and basic business fundamentals; refining the ideas, based on "customer" needs, wants and interests; and gaining a firm understanding of market potential, competition, drivers, trends, and factors. The second part, *Idea2Startup*, is focused on developing a competitive strategy, resolving issues related to intellectual property, building financial models and devising a "go to market" plan. At the conclusion of the two-part course, students have a working proof-of-concept and a realistic business plan for launching their product or service.

2.3.2 Responsive Environments

In Spring 2009, a multidisciplinary class entitled *Responsive Environments* was co-taught by iSchool doctoral candidate Jaime Snyder along with Michael McAllister and Olivia Robinson from VPA Departments of Design and Art, respectively. During this studio-based course, students from across the campus learned how to use video and audio-based information technology to engage with and manipulate their environments through the creation of site-specific art installations. Course material included programming microcontrollers, use of sensors and emitters, and the manipulation of various data and information sources, such as video and sound. Activities provided students with the opportunity to investigate the functional, conceptual and expressive possibilities of weaving together materials, technology and interaction. Studio projects were presented at an exhibition entitled "Wired Space."

2.3.3 Global Enterprise Technology

As part of the iSchool's new *Global Enterprise Technology* minor, Robert Heckman, faculty member and senior associate dean in the iSchool, is using studio-based classroom experiences in order to teach his students about the cultural and social aspects of distributed communication in the corporate world. Through simulations and critiques, he is opening a dialogue with students to foster self-reflection, critical thinking, and hands-on learning. As a result of his studio-based approach, the students are learning the unique skills needed to successfully communicate across continents and time zones by doing, making and responding, rather than just watching.

2.3.4 Design + Virtual Worlds

Design and Virtual Worlds was taught by another senior iSchool faculty member, Associate Dean Jeffrey M. Stanton, along with Michael McAllister, former director of VPA's center for collaborative design (called COLAB). Virtual worlds are graphical computer applications in which users navigate through a 2D or 3D environment with a representation of the self called an avatar. This course provided to students a brief introduction to virtual worlds and the opportunity to test design tools in one virtual world, Second Life. Students also explored the social, educational, entertainment, and commercial uses of virtual worlds. Undergraduate and graduate students from the iSchool, VPA, School of Management, and College of Engineering learned about the operation, limitations, and potential of virtual worlds for various applications. This class met both in person to address technology issues and provide basic tutorials on the use of Second Life, and later in the semester, "in world" at the appointed class time to take guided tours and tutorials.

3. FUTURE PLANS

At the time of this writing, future plans for the *Common Ground* initiative include facilitating interaction between faculty in the art and design fields and their counterparts in the iSchool. This includes:

- Workshops focused on sharing pedagogical techniques and approaches related to studio learning
- Opportunities for interested faculty to identify potential research partnerships
- Support for multidisciplinary proposal development and funded research submissions
- Identification of cross-over topics with possibility for guest lecturing across classes

The purpose of these activities is to provide faculty with the resources to engage students in multidisciplinary practices that span the realms of information, technology, art and design.

In tandem with faculty-centered activities, planned student-focused activities include:

- Dissemination of information about special programs for students hosted by faculty in art and design, including performances, design competitions, field trips and film viewings
- Identification of aspects of department specific capstone programs where design and information students would mutually benefit from collaboration with their peers (such as assignments involving the presentation of information where

information students could contract out the creation of visual communication materials to design students)

- Development of creative and visual thinking courses designed to appeal to a broad range of students

The goal of all of these activities is to provide students with ample opportunity and skills to become broad-spectrum thinkers prepared to engage in a range of multidisciplinary professional activities.

QIC: Query In Context for Educational Collections

Students' demand for rich, multimedia content to be incorporated into their learning process has driven teachers to use online resources. With its accessibility and explosive growth of content, the NSDL repositories are in a prime position to provide the quality material teachers and other knowledge seekers need. Still, teachers are dissatisfied with this resource. They are frustrated with the time-consuming manual effort to retrieve and review each link when searching for the appropriate material. This often leads to settling for "good enough" content. As these collections become larger, the number of hits on a keyword based search will increase. To sustain and increase the utilization of NSDL's quality resources, it is important that a more sophisticated methodology for query and retrieval be developed.

Query In Context for Educational Collections (QIC) is a research project to revolutionize individual search by shifting the burden of information overload from the user to the computer. This is accomplished through context-sensitive text mining methodologies. The major components of QIC's portable unified knowledge discovery system are **context sensitive retrieval**, **semantic query analysis**, and **concept extraction.** Augmenting NSDL's NCore search component with context-sensitive methodologies extends the search engine's capabilities through a modular interface. As context-sensitive text mining research learns more about the variables that support increases in user satisfaction, QIC can be extended to support online searches by minimizing human intervention and increase the relevance of search results.

This research supports sustainability through increased user satisfaction. By organizing the more relevant information first, an NSDL user's time spent in the selection phase of the discovery process is reduced, which makes NSDL's quality repositories and its partners more attractive. Reducing search time should increase content utilization by encouraging repeat use. Finally, higher usage should encourage more new content which in turn will increase visitation frequency.

The NSDL has supported a number of grants analyzing instructor usage of digital libraries. From this research the following key characteristics that influence instructor's search and selection behavior were identified:

- Teachers focus on domain knowledge over pedagogy in most selections.
- Teachers make use of opinion leadership, selecting content from known colleagues or recommendations by their associates.
- There's a higher frequency searching for material to augment a single class than to design a new course. This material is often used in course redesigns.
- The data also suggests a long learning curve (12 months or longer). (Manduca, Iverson, Fox, 2005)

Effective searching, however, remains an issue. McMartin et al. indicated instructors search for the 'perfect' image when preparing for a class but will often "settle for something that is 'good enough'." This trade-off is caused by a "lack of efficient search strategies" and the feeling that "searching for materials can be time consuming" (McMartin, Iverson, Manduca, Wolf, Morgan, 2006). QIC builds on this research to develop its concept

extraction module.

It is expected that as collections become larger, current keyword-based search strategies will exacerbate these frustrations (Recker, M., 2006). Our research utilizes innovative ideas to design efficient information retrieval (IR) and text mining algorithms for large, multimedia libraries.

QIC's goal is to minimize human intervention in the extraction process and reduce the number of contextually inaccurate results displayed. Its unique approach synthesizes user preferences, their situational context, and the informational needs to provide users with results relevant to what they want, rather than presenting 'cookie cutter' answers. This approach will improve user effectiveness and thus satisfaction.

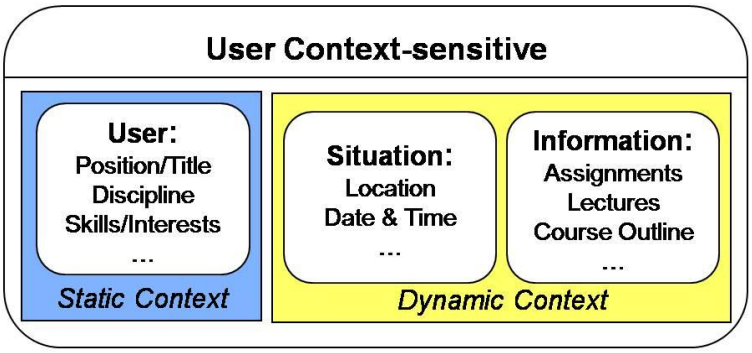


Figure 1: User Context-sensitive Categories

Figure 1 illustrates context information supporting a user's quest for information. A user's static context may include their role (student/teacher), areas of interest (science), and level of education (K-12, undergraduate, etc). Variations of this information are found in login profiles that are standard in an NSDL pathway and other digital libraries. This information does not change frequently and is considered static.

Situation context provides information in terms of where and when. If the data request is made in the middle of a school term it can be inferred to be needed for a class redesign rather than a course redesign. Some studies have shown sentiment or opinion information may be extracted from recommendation systems or blog comments a user may have written in reference to material stored in the repository (Pang, Lillian, 2008 and Hu, Liu, 2004).

The third category captures the user's "information world" – e. g. read documents and visited Web pages – thus reflecting the user's interests. This is an area where the use of text mining techniques has most often been proposed (Mei, Zhai, 2006 and Raymond, 2003 and Fan, Xu, Friedman, 2007).

Certain category variables, when combined with some basic rules may give insight into a user's search to improve context understanding. For example, a *5th grade science teache*r from *Galveston, Texas* (static context variables) types *'wind'* as a search variable on *September 10, 2009* (situation context). We can infer this teacher is not developing a new course because the school year has just started. Most likely they are looking for resources for a lecture or assignment (rather than a test). Because of their location, Galveston Texas, we might rank hurricanes high and specifically give a high ranking to Hurricane Ike, which hit Galveston, Texas on September 1, 2008. While work has been done separately in all three context-sensitive categories, this project proposes the

harmonious incorporation of all three into text mining processes and will augment them with a rules based engine. Figure 2 shows what the output might look like.

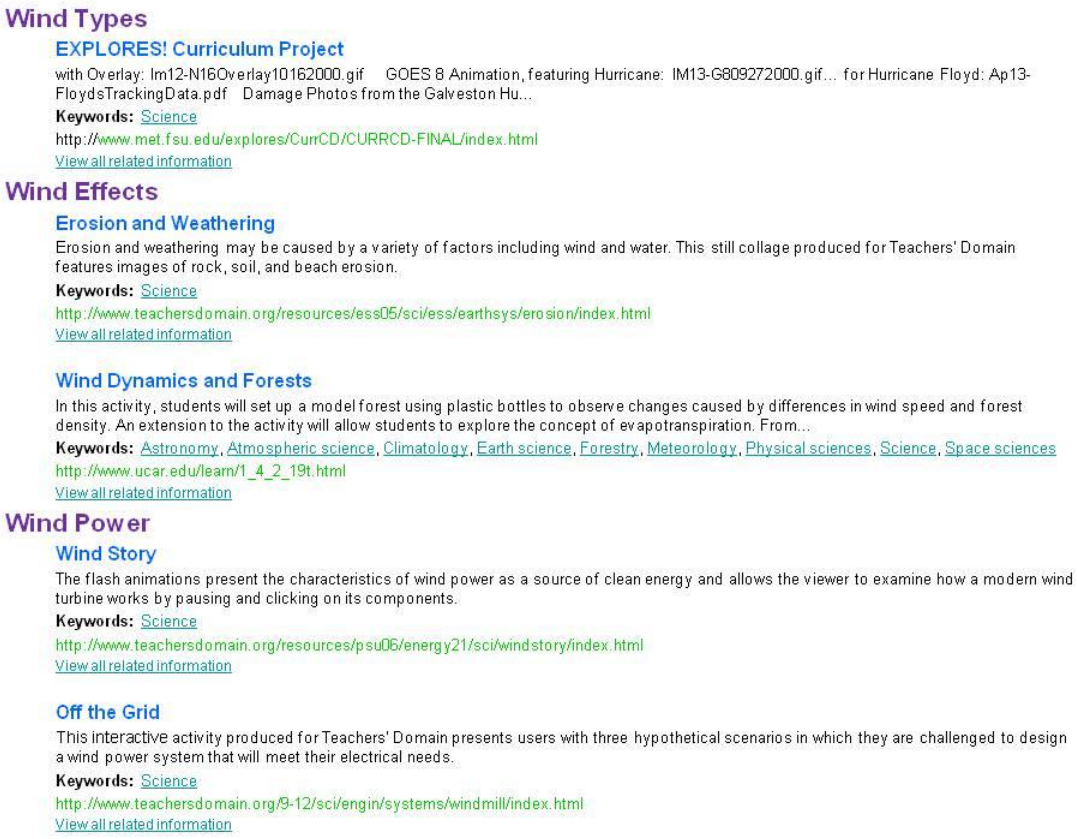


Figure 2: QIC Two-level Context-sensitive Ranked Results

By utilizing data about a user's preferences, search behavior, and information retrieved within the current session, user context-sensitive text mining should provide a more personalized ranked and grouped set of relevant information, thus reducing a user's manual effort in the discovery process. Outliers or results which may lead to accidental discovery or learning will be ranked lower but will not be removed.

Organizing and managing the continuous expansion of digital data is a challenge. QIC helps by integrating digital libraries via a portable platform that supports and improves the discovery process. Techniques to order search results better serves the needs of the users which improves digital library utilization, and, ultimately, encourages the seeking of knowledge and exploration of ideas.

QIC is a starting point to develop a platform portable knowledge discovery system framework that can be tailored to different types of users, content, and digital formats. Evaluations of our results will add to the current research to better understand how educators use digital libraries. This in turn will be a feedback loop to improving extractions results.

REFERENCES

Manduca, C.A., Iverson, E.R., Fox, S.P., McMartin, F. (2005). Influencing User Behavior through Digital Library Design: An Example from the Geosciences, D-Lib, vol 11(5).

McMartin, F., Iverson, E., Manduca, C, Wolf, A., Morgan, G. (2006). Factors Motivating Use of Digital Libraries, JCDL '06 p. 254-255.

Recker, M. (2006). Perspectives on Teachers as Digital Library Users: Consumers, Contributors, and Designers. D-Lib Magazine, 12(9).

Pang , B., Lillian, L. (2008). Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, v.2 n.1-2, p.1-135.

Hu, M. and Liu, B. (2004) Mining Opinion Features in Customer Reviews. Proceedings of Nineteeth National Conference on Artificial Intellgience (AAAI-2004), San Jose, USA.

Fan JW, Xu H, Friedman C. (2007). Classifying biomedical concepts using contextual and lexical features. BMC Bioinformatics. 8: 264.

Can SchoolNet Bridge the Digital Divide in Education in Thailand? Perspectives from Policy Makers to End-users

Wandee Tangsathitkulchai
Department of Library and Information Sciences, University of North Texas, Denton, TX, USA
+940-383-4341

wandee_tan@yahoo.com

Peemasak Angchun
Department of Library and Information Sciences, University of North Texas, Denton, TX, USA
+940-594-5197

surasakandgee@yahoo.com

ABSTRACT

The SchoolNet Thailand, a computer network for Thai schools, was launched by the National Electronics and Computer Technology Center (NECTEC) in 1995 to link secondary schools in Bangkok Metropolitan and provincial areas to the Internet. Under NECTEC's eight-year development of increasing the quality of education, the SchoolNet project has more than 5,000 schools as members. Since then, SchoolNet has become a part of EdNet, the national education network that has the aim to provide Internet connections to schools all over the kingdom. The purposes of SchoolNet Thailand are to develop information technology skills for teachers and to create online lessons covering subjects such as computer, mathematics, science, foreign and Thai languages. As well as develop its website for exchange news and information on activities and training.

SchoolNet is the student-centric educational network that provides equal opportunities for students to get access to the Internet, and the contents and delivery of educational services. Based on NECTEC's evaluation report in 2005, the problem of Internet connection still exists in small secondary schools and some schools that have limited budget. This research proposal investigates the role of SchoolNet in the post-NECTEC era in bridging the digital divide in education among Thai students in terms of access to technology, access to information made available by technology, and access to educators trained in the combination of technology and information into the education experience. The research question will find out how the three accesses can satisfy educational needs both in terms of Internet connection, the contents and delivery of educational services. We will conduct survey research. The sampling frame is secondary schools that are members of the SchoolNet project. Participants will be secondary school principals, teachers, and students. We will use the stratified sampling method to get participants from each regional area to take part in this study. Moreover, this study uses two techniques to collect data consisting of three sets of survey questionnaires. The Likert scale is used to measure satisfaction level of the content and delivery services. We use descriptive statistics such as frequency, percentage rates, arithmetic mean, and standard deviation to analyze data. The data are analyzed using the program SPSS for Windows. We also interview policy makers of the Ministry of Education of Thailand for their future plan and visions on this network.

A Classification of Agents and Entities Influencing Law Enforcement Agencies in the United States

Joseph Treglia
Syracuse University
245 Hinds Hall
Syracuse, NY
1 (315) 443-2911
jvtregli@syr.edu

ABSTRACT
This work introduces a typology for classifying the entities and groups that influence law enforcement agencies in the United States. The purpose here is to begin the process of creating a more formal taxonomy to better understand the types of influences that bear upon law enforcement agencies. Through that work we will better understand policy and decision making in this distinct arena and improve law enforcement operations. Using a soft systems approach this initial work involved a broad search of entities that may influence law enforcement agencies. The search involved extensive internet keyword searches, reviews of paper publications and field observation. In consultation with other researchers the list was reviewed and analyzed for classification. Seven distinctive categories were identified; Law Enforcement, Government, Quasi-Government, Associations, Vendors, Media and People. The resulting typology may be used as a basis for further research in the area of law enforcement organizational behavior, policy development and program implementation. This work will lead to a more formal taxonomy for understanding these relationships.

Categories and Subject Descriptors
H.1.0 [Information Systems] Models and Principles - General
K.4.1 [Computers and Society] Public Policy Issues – Use.
K.6.1 [Management of Computing and Information Systems] Project and People Management – Strategic Information Systems Planning.

General Terms
Human Factors, Standardization, Theory.

Keywords
Influence, Classification, Stakeholders, Typology, Law Enforcement, Collaboration, Organization.

1. INTRODUCTION
"Policing in most societies exists in a state of 'dynamic tension' between forces that tend to isolate it and those that tend to integrate its functioning with other social structures" (Clark, 1965). In truth law enforcement agencies engage with an extensive network of formal and informal contacts both on and off the job. It is important to understand these contacts and their potential influence in agency operations and activity yet little work has been done in this particular area (McKelvey, 1975; Mills & Newton Margulies, 1980; Pugh, Hickson, & Hinings, 1969; Reiss, 1992). Our interest is in improving the way law enforcement agencies fulfill their individual missions by creating a better understanding of the factors that influence their decision making and operation and the environment they work in.

2. PURPOSE AND DEFINITIONS
We wish to improve law enforcement operations. Here we try for the first time to comprehensively identify and classify stakeholders that influence law enforcement agencies. Having such a typology will stimulate thinking and serve as a basis for theory development in this area. Using a soft systems methodology we introduce a comprehensive schema for categorizing law enforcement stakeholders. Law enforcement organizations have distinct member non-member identification; a minimum requirement for formal organizations (Aldrich & Herker, 1977). Through this work we will identify significant actors or entities that have influence within and across the boundaries in this environment and assess their relative impacts.

For our purposes law enforcement agency is defined as one whose mission involves the enforcement of laws and whose personnel are authorized to make arrests and carry firearms. There are many agencies who are involved in investigations that do not make arrests or carry firearms and these agencies are not included here so that we may proceed with an identifiable group consistently across federal and local levels. These agencies will be listed where they interact with or influence law enforcement agencies. An example is the case of the Child Protective Service in New York State; these investigators of child abuse have investigative responsibility but do not carry firearms and they interact routinely with law enforcement. Some federal or other law enforcement agencies may not have the granted authority to carry firearms in all states. In these cases the agency qualifies as a law enforcement agency if the agency has arrest power and authority to carry firearms in even one location or jurisdiction.

3. PROBLEMS
There is significant overlap in areas of jurisdiction and responsibility and there are a variety of ways in which individual agencies choose to organize themselves. We identify agencies and entities that influence or impact law enforcement decision making and activity at the organizational level. Influencers include those that interact internationally, nationally, by state, regionally or locally. Influencing entities or agencies may have influence on agencies that do not follow geographical or political boundaries. Products or services may be used by agencies in a variety of jurisdictions without geographic or political linkage. There are cases where personal contact or affiliation will have a strong impact, such as the case where an executive member has a close relative in some service industry and this relationship causes that

company to have closer ties to the police agency than they may have if there were no outside connection. For this part of the study we include the vendor agency as having influence or being a stakeholder to that law enforcement agency because there is some interest being served at the agency level. If it were the case that the close family relative had a business that had nothing to do with the law enforcement agency interest it is unlikely that even with the familial tie that there would be any collaboration or influence. This is to say that it is valid to investigate agencies and entities which influence or are stakeholders to law enforcement even where there may be outside influences or interests that exist.

4. METHODOLOGY

Data collection for law enforcement agencies and entities that may be associated with them came through a variety of sources. Sources included formal publications, web searches and article reviews such as in the following table:

Table 2: Sources

Source	Retrieved from
Federal Bureau of Investigation Publications - FBI Law Enforcement Bulletin. (n.d.).	August 24, 2009, from http://fbi.gov/publications/leb/leb.htm
NAPO : National Association of Police Organizations.	August 24, 2009, from http://www.napo.org/
Open Directory - Society: Law: Law Enforcement: Organizations.	August 24, 2009, from http://www.dmoz.org/Society/Law/Law_Enforcement/Organizations/
Police & Law Enforcement - Officer.com: Police News, Forums, Links & More for Police Officers, Law Enforcement, Corrections, Sheriffs & More.	August 24, 2009, from http://www.officer.com/
Police Magazines and Publications.	August 24, 2009, from http://mainesecurity.com/Police_Magazines_and_Publications.htm
Police: Organization and Management - The American System Of Policing, Variation In Style And Structure, Managing Police Organizations, Information Technologies And The Police.	August 24, 2009, from http://law.jrank.org/pages/1675/Police-Organization-Management.html
U.S. Government Bookstore: Law Enforcement Publications from the Federal Government.	August 24, 2009, from http://bookstore.gpo.gov/collections/crime-prevention.jsp

Sources including Law Enforcement Online (LEO), Google and DMOZ, the online open directory project at http://dmoz.org, and Wikipedia were used to identify agencies, businesses, stakeholders and associates. Results from these searches resulted in more than 6,000 separate entries of varying types. A review of the results was done to identity a schema for organizing the results into the most limited number of categories that would place them into logical categories with exclusivity. In consultation with two other researchers the list was reviewed and analyzed for classification. Through this continuing work we intend to refine and formalize the identified categories that make up this schema for understanding agents and entities influencing law enforcement agencies in the United States.

5. RESULTING TYPOLOGY

The typology proposed includes the major categories of: Law enforcement, Government, Quasi-Government, Media, Associations, Vendors, and People. These identified categories are described in the following table:

Table 1: Categories of Agents and Entities

Category	Description or included entities
Law Enforcement	Outside US Agencies, Special Jurisdiction, Federal, Tribal, State, County, City, Town, Village
Media	Local News Media, National News Media, Entertainment Media - News Like, Entertainment Media - Shows, Broadcast TV, Blogs, Online News Media, Online - i.e. YouTube, Web Sites
Government	Village Government, Town Government, City Government, County Level Government, State Level Government, Federal Level Government, Tribal Governance, School District, Fire District, Other Special Government Unit
Quasi-Government	Power Company, Water Authority, Telecommunications Providers, Postal Service, Rail Service
Associations	Police Benevolent Associations, Unions, Professional Associations Agency, Professional Associations Individual, Collaborative Partnership Associations, Supportive , Lobbies Citizen Advocacy groups, Religious, Political
Vendors	General Businesses & Service Providers: Training and Research, Equipment and Tools, Hardware Providers Communications , Software Providers Data and Communication, Supportive Services Health, Maintenance and Contract Services, Consultants, Insurance, Retirement
People	Citizens, Family, Friends, Bad Guys - Criminal Element

5. CONCLUSION

The greater goal is to improve the functioning of law enforcement operations in this country. Seven distinctive categories are

identified as a format for analysis and discussion. The typology may be used as a basis for further research in the area of law enforcement organizational behavior, policy development and program implementation. This work should lead to a more formal taxonomy for understanding these relationships, aid in theory development and stimulate thinking in this area.

6. BIBLIOGRAPHY

Aldrich, H., & Herker, D. (1977). Boundary Spanning Roles and Organization Structure. *The Academy of Management Review*, *2*(2), 217-230.

Clark, J. P. (1965). Isolation of the Police: A Comparison of the British and American Situations. *Journal of Criminal Law, Criminology and Police Science*, *56*, 307.

McKelvey, B. (1975). Guidelines for the Empirical Classification of Organizations. *Administrative Science Quarterly*, *20*(4), 509-525.

Mills, P. K., & Newton Margulies. (1980). Toward a Core Typology of Service Organizations. *The Academy of Management Review*, *5*(2), 255-265.

Pugh, D. S., Hickson, D. J., & Hinings, C. R. (1969). An Empirical Taxonomy of Structures of Work Organizations. *Administrative Science Quarterly*, *14*(1), 115-126.

Reiss, A. J. (1992). Police Organization in the Twentieth Century. *Crime and Justice*, *15*, 51-97.

Modeling Staff Behavior in the Production of Information Products

Cameron Tuai
School of Library and Information Science
Indiana University Bloomington
1320 E10th St. LI 011
(812) 855 2018
ctuai@indiana.edul

ABSTRACT

Improved understanding of staff behavior in organizations whose primary product or competitive advantage lies in the production of information connects with two trends in the information industry: (1) the growing import of bundled, or meta-information, in corporate sales strategy, as exemplified in the information sales strategies of companies as EBay, Expedia, or Amazon (2) a shift in the competitive information organization's landscape from using expertise in developing, deploying, and managing ICTs to create barrier to entry, to ICTs as commodities [4]. These two trends suggest a movement away from the ICT as the central artifact and towards that of a contextual factor within larger information production process. Further, under such conditions, the behavior of staff involved in the production of information products becomes increasingly important in achieving a competitive advantage. My research posits that an information organization capacity to compete partially lies in its ability to foster staff behaviors that produce optimal usage of ICTs in acquiring, organizing, and disseminating information. This research contributes to the field of information studies by (a) developing the models and methodologies for examining these types behavior; (b) identifying the patterned variations that differentiate these types of organizations from others; and (c) creating theories of causation regarding the relationship among organizational structures, ICTs, and staff behaviors toward production of information products. The goal is to model the fit between staff behavior, ICTs, and organizational actions that best optimizes the production of information products.

My theoretical approach draws from both information studies (IS) and organizational theory. This approach allows me to keep one eye on the IT artifact and the other on other factors at work within the organization. Support for IS research that moves away from the information technology (IT) artifact as the central focus [3] is provided by both Galliers [5], who sees the future of IS as a shift from IT as the central artifact to that of people/information, and DeSanctis [4] who suggests that natural evolution of IS is away from IT as artifact, and toward IT as a human or organizational challenge. Orlikowski and Barley [9] write that both organizational theory and IS offer certain advantages to the study of IS problems. They note that drawing from the field of information studies provides a means of understanding technology as both a social and physical artifact. This conception of technology allows for a "more nuanced appreciation for why and how the material properties of technologies matter" and the development of "better images of how forms of organizing emerge as human action weaves itself around a technology's constraints and affordances" [9]. The benefit of organizational theory, on the other hand, lies in its ability to provide the broader framework for discovering regularities, general principles, and causal relationships. Drawing from both disciplines also allows me to build upon organizational theory's rich body of empirical research, while having opportunities to explore research anomalies using IS approaches such as those employed in social and organizational informatics.

My research context is information service units formed through a partnership between academic libraries and campus computing. The growing use of partnerships to create information products is documented in both the library and information science literature as well as in areas of the IS literature such as IT governance. Within the context of academic libraries, these types of units are often referred to as information commons, learning commons, or research commons. These hallmark of these units is the combination of librarians and technologists within an ICT-rich environment in order to facilitate customer knowledge creation [1]. Studying the library form of this information production partnership affords both depth of the relevant professional literature and a large number of units in operation (I have identified approximately 110 for my initial sample). The literature on the collaboratively based information services reveals a number of issues that pique my curiosity. To begin with, the creation of knowledge products within these units suggests a high level of behavioral integration between librarians and technologists. Structural contingency theory posits that managers can best achieve this level of integration through horizontal structures. Although the case literature frequently describes horizontal structures, there is little evidence of the presence of services that require such resource-intensive approaches. What factors, other than high degrees of integration, are causing managers to create horizontal structures? How are these units fulfilling the promise of knowledge creation if not through services that require high levels of integration? Are the commons fulfilling their promise to provide the knowledge creation products? If they are not, what holds them back?

To explore some of the organizational anomalies found within the professional information services literature, I use contingency theory to examine the relationships between people, organizations, and IT; factors that March describes as the interface of research in information systems [8]. I measure staff behavior using the degree of behavioral interdependence present in the production of the unit's information product. Structural

contingency theory posits interdependence as the explanatory variable for structural coordination. As such, I use coordination to represent organization forces. The contingency expectation of a positive relationship between interdependence and coordination forms the starting point for the development of a model of staff behavior in information producing organizations. Successful management of highly *interdependent* behaviors in ICT usage can lead to innovation in the creation of information products. The successful management of highly *independent* behaviors in ICT usage can lead to efficiencies in information operations. This model contributes to the IS literature by offering additional explanations for variances within information systems structures. For example, Barley's [2] seminal article on the effect of CT scanners in radiology departments concludes that while the introduction of technological uncertainty resulted in decentralization, the degree of decentralization depends on the specific historical process in which they are embedded. Reviewing this article in terms of a contingency perspective on interdependence offers an alternative explanation many of the differences found in Barley's descriptions of specific historical processes.

Turning to the last concept within March's [9] conception of an information system, the measurement of IT. This factor presents a measurement challenge in situations where the production of information involves partnerships. Within the context of my research setting of collaborative information service units, the two partners are generally composed of highly differentiated groups of information professionals. Typically each information profession brings with it histories that differentiate staff in regard to their approach to such issues as service levels, appropriate use, costs and benefits, or goals. Further, each profession focuses on different types of ICTs within the information service point, which is unto itself an ICT system Given this situation, Orlikowski and Barley [9] quite convincingly argue that IS's conception of technology is superior to contingency theory's conception of technology in explaining the process of organizing within an ICT-intensive context. The conundrum is that the IS approach to measuring technology is epistemologically incompatible with my desire to build a contingency theory based causal model. The compromise was to use Lawrence and Lorsch's {, 1967 #712} instrument for measuring behavioral differentiation. This instrument is at best a loose proxy for an IS conception of technology, but its measurement of behavior in terms of goal, time, and interpersonal orientation should capture part of the underlying forces that are eventually expressed through IS's conception of technology as both a social and physical artifact.

My research into integrated information service unit seeks to build upon prior work that has largely failed to confirm expected contingency theory relationships. Analysis of these studies suggests that this problem results from researchers underestimating the complexity of the information production context in terms of the relationship between interdependence and coordination and the potential moderating effect of technology derived behavioral differentiation [7,10,11]. This underestimation leads to large units of analysis, such as divisions, that are internally heterogeneous or externally homogeneous, thus confounding efforts to confirm expected relationships. My research resolves this problem through two different methodological designs. First, my focus on integrated information service units represents a smaller unit of analysis thus decreasing the presence of extraneous variables. Second, I accept the IS definition of technology as a complex social and physical artifact and as such substitute interdependence and behavioral differentiation as the causal variables within information service units.

REFERENCES

[1] Bailey, D.R. and Tierney, B. Transforming library service through information commons : case studies for the digital age. American Library Association, Chicago, 2008.

[2] Barley, S.R. Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments. Administrative Science Quarterly, 31 (1). 78-108.

[3] Benbasat, I. and Zmud, R.W. The identity crisis within the IS discipline: Defining and communicating the discipline's core properties. MIS Quarterly, 27 (2). 183-194.

[4] DeSanctis, G. The Social Life of Information Systems Research: A Response to Benbasat and Zmud's Call for Returning to the IT Artifact. Journal of the Association for Information Systems, 4 (1). 16.

[5] Galliers, R.D. Change as crisis or growth? Toward a trans-disciplinary view of information systems as a field of study: A response to Benbasat and Zmud's call for returning to the IT artifact. Journal of the Association for Information Systems, 4 (6). 337-351.

[6] Lawrence, P.R. and Lorsch, J.W. Differentiation and integration in complex organizations. Administrative Science Quarterly, 12 (1). 1-47.

[7] Lynch, B.P. An Empirical Assessment of Perrow. *Administrative Science Quarterly*, *19* (3). 338-356.

[8] March, S.T. Designing design science. Information systems: The state of the field, John Wiley & Sons. 338-344.

[9] Orlikowski, W.J. and Barley, S.R. Technology and institutions: what can research on information technology and research on organizations learn from each other? MIS Quarterly. 145-165.

[10] Weng, H. A. A contingency approach to explore the relationship among structure, technology, and performance in academic library departments. in Williams, D.E. and Garten, E.D. eds. *Advances in Library Administration and Organization*, JAI Press, Connecticut, 1997, 249-317.

[11] Vorwerk, R.J. The environmental demands and organizational states of two academic libraries *School of Library and Information Science*, Indiana University, Bloomington, 1970

Using Paper Maps for Geospatial Data Collection

Sarah Van Wart
UC Berkeley School of Information
University of California Berkeley
Berkeley, CA 94709
vanwars@ischool.berkeley.edu

Michael Manoochehri
UC Berkeley School of Information
University of California Berkeley
Berkeley, CA 94709
michael@ischool.berkeley.edu

Nathan Gandomi
UC Berkeley School of Information
University of California Berkeley
Berkeley, CA 94709
ngandomi@ischool.berkeley.edu

ABSTRACT
Geo-spatial information is important for a variety of decision-making, planning, marketing, and development activities. Yet, in the developing world, there are significant challenges to collecting this data, including poor network infrastructure, the lack of availability and trained operators of hand-held devices, and the difficulty in finding specialists who can effectively organize spatial data. This project explores the possibility of using a simpler GIS data collection methodology – annotating paper maps – in order to both simplify the data collection process, and capture more qualitative data. Through interviews, prototyping, and testing, we hope to design a viable proof of concept that can both support existing GIS data collection efforts, and lower the barrier to entry for additional organizations to collect geo-spatial data and conducting analyses.

Categories and Subject Descriptors
H.2.8 [**Database Applications**]: Spatial databases and GIS, Image databases.
H.3.5 [**Online Information Services**]: Data sharing, Web-based services.
H.4.2. [**Types of Systems**]: Decision Support.
H.5.1. [**Multimedia Information Systems**]: Hypertext navigation and maps.
H.5.2. [**User Interfaces**]: Graphical User Interfaces (GUI), Prototyping.

General Terms: Design, Human Factors, Experimentation.

Keywords: ICTD, Geospatial Data, Participatory Mapping, Rural Development.

1. INTRODUCTION
During the summer of 2009, the University of California at Berkeley, with funding from the Gates Foundation, launched a research project to investigate the potential for information and communications technology to help improve the economic conditions of smallholder farmers in Ethiopia and Uganda. Teams of graduate students, with support from local NGOs, conducted qualitative research on the ways in which farmers, community groups, and development organizations use available communications technology. A common observation was that organizations working in rural development often encountered challenges in obtaining current and accurate spatially referenced data about the communities they wish to serve.

1.1 Observations in the Field
While in Ethiopia and Uganda, we made several key observations that have informed this project. First, many of the development organizations we observed did not have sufficient capacity to collect data in the field or process it using their existing GIS software. Second, we observed that there was an unreliable / non-existent electricity and cellular network coverage in many of the rural field sites we visited, which reduced the usefulness of collecting data using networked digital equipment. Especially in Ethiopia, we found that a lack of reliable mobile telecommunications infrastructure meant that data collected in the field would have to be transported by hand, or by truck, to a central office for processing. Third, we observed that many organizations used participatory mapping to facilitate dialogue and to help reach consensus on a variety of development topics, such as irrigation, crop production, and forest management. Though the resulting map artifacts contained a wealth of information, these maps were seldom used by planners or incorporated into organizations' larger GIS systems. Fourth, we observed extension agents recording a variety of data about farmer yields and crop production. It was unclear that these forms were ever propagated up to development organizations for analysis.

1.2 Participatory Mapping and Community
For years, organizations have been using participatory mapping techniques to facilitate dialogue and to help reach consensus on a variety of development topics, such as irrigation, crop production, and environmental management [1]. This process usually involves a facilitator, who uses a script or template to ask community members a series of questions [2]. From this dialog, a map is drawn – often on paper, but occasionally drawn on the ground or in the sand – where borders are delineated, problem areas are highlighted, and important issues are brought to light from the perspective of the community members.

Geospatial data can be gleaned from these community-generated maps and provide development organizations with valuable information and perspective that is difficult to gain using other data collection techniques, such as interviews and surveys. Data provided by different community members can be aggregated using the same map, providing an efficient view of complex social behavior. Organizations are able to combine data collected in the field with other data provided by outside sources, to help improve knowledge and decision-making. The maps created can also be used as artifacts to help facilitate dialogue between community members and development workers.

1.3 Project Goals
Given these limitations on human capital and telecommunications infrastructure, fieldworkers' affinity for paper-based data collection, and the potential for hand-drawn maps to reveal community values, perceptions, and tacit knowledge, our project team will explore alternate methods for geo-spatial data collection and place-based communication. We wanted to create a simplified GIS data collection system, which was also capable of

handling hand-drawn maps, and incorporating more actors to participate in data collection.

2. RELATED WORK

There are a number of geo-spatial data collection and dissemination initiatives. Regarding open platforms that accept user-generated geo-spatial content, Google Map Maker and Open Street Map [3] are two of the most prominent, both of which have specific initiatives geared toward digitizing maps in the developing world. For example, in November 2009, residents of Kibera, Kenya's largest slum, used Open Street map and GPS units to annotate what had previously been a blank spot on a map. There are also a number of open source mapping tools, such as Modest Maps and Open Layers, and publicly accessible APIs, such as the Google Maps API and the W3C Geolocation API, that allow users to consume, display, and query existing geographic data. There is also an emerging field of Participatory GIS (PGIS) in which publicly generated data has been incorporated into GIS systems in a variety of ways, in Pittsburg [4] [5], the Tigray region in Ethiopia [6], and Chicago [7], to name a few.

3. PROPOSED WORK

Guided by our field research in Ethiopia and Uganda over the summer, we plan to explore how low-tech data collection and high-tech data processing might be (1) a cheaper, faster, and richer way to gather data, and (2) a medium to enhance the communication of more complex, qualitative ideas between parties. That is, by creating a prototype that leverages paper-based maps, image recognition tools, and GIS, in conjunction with a set of organizational processes, we will demonstrate that paper-based, geo-spatial information gathering can be efficiently streamlined into modern data collection and communication systems in new ways.

We will start with the existing Walking Papers [8] code base, which is a system that allows a user to print a barcode-indexed paper map from a web-based system, annotate this paper map manually, scan it, and incorporate it back into the system using image recognition software. We will then look at how the existing Walking Papers UI might be adapted and supplemented by process to address certain data collection and communication challenges. Questions to explore include: how might fax or postal `mail be used as a data transmission alternative? Could using ink-based stamps with particular shapes or ordinary pens with certain characteristics assist computer vision technologies in automatic feature extraction tasks? For which types of data collection could paper maps work? In which scenarios would hand-drawn process- or perceptual- maps interface with GIS systems in useful ways?

With respect to data collection, we will explore specific scenarios in which a paper-based data collection approach might be preferable to using GPS units or sensor-enhanced mobile devices. Though a paper-based geospatial data collection strategy might not be sufficient for all data collection initiatives, it could certainly be a viable substitute in particularly resource-constrained environments or in cases where extremely time-sensitive data needs to be collected on a massive scale. We will also explore the processes that would be needed to ensure a successful data collection effort.

With respect to communication, hand-annotated paper maps can take advantage of affordances that are absent from most spatially oriented hand-held devices. On a map, important areas can be circled, processes and directionality can be expressed, and multiple parties can collaborate together to express a single idea. These more qualitative, information-rich expressions of ideas are often lost in more formal GIS systems, but understanding community perceptions is critical to the success of projects. Our prototype will explore how qualitative spatial expressions of communities and individuals could interface with formal GIS datasets (infrastructure, natural features, etc.), using location as the unifying feature. Again, we will explore the process and technology elements needed to best facilitate this type of communication.

3.1 Prototype Development and Testing

In light of the exploratory research we have already conducted over the summer, we will now begin to validate the usefulness of our prototype concept with potential users of the system. Though it would be ideal to ultimately test the viability of a paper map data collection initiative in East Africa, in the meantime we plan to test our prototype with a local organization that runs an environmental science program for low-income youth in the Bay Area. The science organization already uses a participatory learning approach and coordinates a variety of qualitative and quantitative data collection exercises, where students collect location-based information about their communities, including air and water quality data as well as perceptions about the safety of their neighborhood. Our research team plans to work with the organization to support its more qualitative data collection efforts. Collective brainstorming is currently underway.

3.2 Evaluation

We propose to evaluate the effectiveness of our prototype in two ways. For the data collection portion of the project, we will compare the costs (equipment, training, etc.) and amount of time needed to collect and aggregate geospatial data between hand-drawn and device-generated methods. Secondly, for the qualitative, communication portion of the project, we will conduct a series of interviews to try and determine how useful geo-referenced, hand-drawn maps are to organizations.

4. ACKNOWLEDGMENTS

We want to thank George Scharffenberger and Dr. Tapan Parikh, for guiding us in our research and enabling our research trip to East Africa.

5. REFERENCES

[1] Shankar Aswani and Matthew Lauer, *Incorporating Fishermen's Local Knowledge and Behavior into Geographical Information Systems (GIS) for Designing Marine Projected Areas in Oceania. Human Organization*, 65 (1). 81-102.

[2] King, S., Conley, M., Latimer, B., and Ferrari, D. *Co-Design: A process of design participation.* Van Nostrand Reinhold Company, 1989.

[3] M. Haklay and P. Weber, *OpenStreetMap: user-generated street maps*, Article, October 2008.

[4] Vajjhala, S.P. *"Ground Truthing" Policy: Using Participatory Map-Making to Connect Citizens and Decision Makers.* Resources-Washington DC, 162 (14).

[5] Aynekulu, E. and Wubneh, W. *Monitoring and Evaluating Land Use/Land Cover Change Using Participatory Geographic Information System (PGIS) Tools: a Case Study of Begasheka Watershed, Tigray, Ethiopia.* EJISDC 25, (3). 1-10.

[6] Al-Kodmany, K. *Using visualization techniques for enhancing public participation in planning and design: process, implementation, and evaluation. Landscape and Urban Planning*, 45 (1). 37-45.

[7] Vajjhala, S.P. *Integrating GIS and participatory mapping in community development planning*. in proceedings of the *Twenty-Fifth annual ESRI User Conference*, (San Diego, CA, 2005), http://gis.esri.com/library/userconf/proc05/papers/pap1622.pdf, Citeseer.

[8] Migurski, Michael. *Walking Papers*. Retrieved March 15, 2005, from *Walking Papers*: http://walking-papers.org.

[9] Vajjhala, S.P. *Integrating GIS and participatory mapping in community development planning*. in proceedings of the *Twenty-Fifth annual ESRI User Conference*, (San Diego, CA, 2005), http://gis.esri.com/library/userconf/proc05/papers/pap1622.pdf, Citeseer.

The Role of Public Libraries in Society:

A Case Study from a Poor Suburb of
Windhoek, Namibia.

By Sarah M. Webb, Syracuse University

ABSTRACT

Introduction:

This research focused on empirical evidence to clarify the role of the public library in society. Public libraries are a particular type of information provision, where information is thought of as a public good. Ideas of information as a public good lead to discussions of the role of information and information provision in democratic societies. The researcher believes that the more we know about the roles of various information providers in society, the better we will be able to make policy for the provision of information, whether it is through market forces or through government sponsored provision.

Background:

The UNESCO Public Library Manifesto [1] defines the public library as an organization, which helps create a democratic, equal and peaceful society. This definition helps to justify the creation and cost of public libraries for societies around the world. Buschman (2007)[2] and Kabamba (2008)[3] both open the question of whether the libraries actually live up to the promises of the Manifesto. Buschman notes that there is little empirical evidence to support the link between public libraries and democracy. Indeed, many non-democratic countries (e.g. the Soviet Union and Nazi Germany) had excellent public library systems.

Kabamba (2008) notes that the particular services offered by many public libraries in African countries limit the ability of the library to realize the promises of the Manifesto. He states that the libraries focus on providing a place for children to learn and grow, and do not provide enough services for adults.

A case study was carried out in a suburb of Windhoek, Namibia – Katutura. This site was chosen because Namibia is a newly independent nation, which has made public libraries available to all citizens within the last 25 years. This means that the role of the library is being constructed within society. As well, Katutura was the former black township of Windhoek and is still a poor neighborhood. Most libraries in Africa are urban, often situated in areas, which are difficult for the poor to use. By situating the study in a lower class neighborhood, it was possible to see if the library could help equalize society.

Research Questions:

Many librarians, including the researcher, believe in the promise of public libraries to create peaceful, democratic and equal societies. Empirical evidence is needed, however, to determine if they do. As Buschman points out, determining how libraries help with the creation and continuation of democratic societies is not easy. Thus, this research sought data that might elucidate the role of the public library in society.

This research question is not easy to answer on its own, so two sub-questions help clarify the role of the library in society. As Wiegand and Bertot (2003)[4] noted much has been written on the user in the life of information systems and libraries, but little has been written on the role of the library or information system in the life of the user. The first sub-question, therefore, is *"What is the role of the library in the lives of a person?"*

Many researchers have noted (e.g. Mchombu, 1982)[5] that public libraries in Africa are often built to serve the educated elite. For that reason, uncovering who uses the library and gaining an understanding of the identity issues surrounding library use also seemed important. This led to the second sub-question, *"Who uses the library?"*

Methods:

The case study was carried out following Burawoy's Extended Case Method[6]. Most of the data was collected through participant observation, with interviews and document analysis for triangulation.

Results:

Who uses the library?

Interviews with people in the library and people in the neighborhood around the library revealed that the library was for students and learners. This emphasis on education and library use as a key to life success was clear. People using the library were assumed to want or to have success in school.

Individual Role:

The identity issues around use of the library reflect the role of the library for individuals, which is mostly one of education support. Students and learners, people in university programs and secondary schools, need a place to study, resources to aid their study, and the motivation that comes from seeing other people working hard. Younger learners often use the library as a space to do homework and to learn informally.

For the younger generation, the library is seen as a key element of their educational success and their educational success is seen as a key element to their potential life success. The secondary school students and university students in the library, in general, are hard working and optimistic about their country and their life-goals.

Adults using the library are either using it to help them find a job or using it to read the newspaper. For the most part, men read the newspaper in the library, and seemed to have the leisure time to spend hours in the library reading the newspaper or magazines or books. Older women were not seen doing these things, although some young women would.

The libraries partly played a role in people's lives as a source of information, but their more important role seemed to be as a place for knowledge creation. People studied to create knowledge about their school subjects, and people created information documents to help them find employment. The library provided a quiet space to think and the tools, photocopier and computer, to make employment documents.

Societal Role:

One way to ascertain the societal role of the library would be to project the micro-role of the library to the macro-level. The library is helping students and learners attain success at school. Therefore, the role of the library in society is to help with formal education. One could also say it plays a role in making newspapers available, and the provision of newspapers and a free press can be tied to the building of a democratic society. The role is also to provide people with the space and tools to create knowledge and improve their life chances.

The role of the library in society is hampered, however, by the fact that the libraries are few in number and small. Thus, defining the role by those who use the library, ignores the fact that most people do not live close enough to a library to be able to make use of what it may offer. Since Namibia became an independent nation, the number of public libraries has increased from 23 to over 60.

The government of Namibia has a strong stated desire for a knowledge society and an economy based on knowledge work. Statements in the Namibia Vision 2030 for the society and in various development plans all speak to this desire. This desire is reflected in the notion that learning and school success will help with life-success. The notion that each individual's success at school will make for a more successful and developed nation is an ideal heard repeatedly in interviews, conversations and government speeches.

The library administration tries to make the case that libraries are a key part of this strategy - in order to create a knowledge society, people will need access to information. Although the belief that education is a key to individual and national success, this has not always translated into more library programming, although it may be part of the reason that more libraries exist today. As in other countries, however, advocating for libraries and ensuring good budgets for libraries is a difficult job.

Conclusion:

In conclusion, my findings indicate that for libraries to have an impact on society, everyone needs to have access to them. Even though Namibia has tripled the number of public libraries, they still have so few that most people in the country do not have access to public libraries. The desire for a knowledge economy is strong, however, and libraries could be part of the achievement of this goal. Many documents attempt to position the libraries to do this, but limited budgets have kept libraries from realizing their potential.

At the same time, when libraries are available they play a substantial role in the lives of the people who use them. Many people use the library every day or multiple times in the week. These people primarily use the library to support their formal education. A small but persistent number of young adults also use the library to read the newspaper and help with job searching. In this way the libraries are living up to some of the promises of the Manifesto.

1. REFERENCES

[1] UNESCO, & IFLA. (1994). UNESCO Public Library Manifesto. Retrieved February 2, 2008, from http://www.unesco.org/webworld/libraries/manifestos/libraman.html

[2]Buschman, J. (2007). Democratic Theory in Library Information Science: Toward an Emendation. *Journal of the American Society for Information Science, 58*(10), 1483-1496.

[3]Kabamba, J. M. (2008). *Libraries re-loaded in service of the marginalized.* Paper presented at the Standing Conference of East, Central, and Southern African Library and Information Associations, Lusaka, Zambia.

[4] Wiegand, W., & Bertot, J. C. (2003). New Directions in LQ's research and editorial philosophy. *Library Quarterly, 73*(3), v-ix.

[5] Mchombu, K. (1982). On the Librarianship of Poverty. *Libri, 32*(3), 241-250.

[6] Burawoy, M. (1998). The Extended Case Method. *Sociological Theory, 16*(1), 4-33.

Which Life-Cycle?
Data Curation and Records Management Education

Nicholas Weber
University of Illinois, Champaign- Urbana
710 W. Church St.
Champaign, IL 61820 USA
+1 312 330 1187
nmweber@illinois.com

Abstract

In this poster I will present the two life-cycle models of records management and data curation. I will use these figures to compare and contrast the two disciplines, and ultimately suggest where they might benefit from future educational collaborations.

Keywords

Records Managemet. Data Curation, Collaboration, Library and Information Science, Gradaute Education.

Introduction

In predicting the data deluge[1] researchers and theorists have made numerous declarations of what functions data curators can and should perform in life-cycle management. But relatively little attention has been paid to the theory and practice of electronic records management, which encompasses many of the same actions . Are data curators unnecessarily re-inventing life-cycle models or does the very scale of this data require curators to contemplate new theorizations of management services? This poster will explore the similarities of life-cycle management in data curation [2] and records management [3]in order to show similarities that exist within the two disciplines. Ultimately, this poster hopes to explain how the two might collaborate in future educational efforts and why this is an especially important time for harnessing these collaborations

Body

Traditional principles of records management have focused on record authenticity, integrity, provenance and long term value [4]. These considerations are also at the core of what data curators must consider in similar practices of appraisal and selection [5]. The two disciplines overlap in numerous ways, but none as immediate as the graphic depiction of the "life-cycle" model that both domains widely distribute. Practitioners in both fields refer to their own particular life cycle construction in order to explain the workflow of their actions, but perhaps its time they start showing these models to one another.

Additionally, the vast increase in the amount of modern data also comes with new complications for both electronic records managers and data curators. Electronic records managers are encountering data in formats and at a scale previously unprecedented in the business environment, while data curators are quickly being saddled with legislative mandates requiring they make publicly funded data accessible to society as a whole.

Each of these issues are positive outcomes of a changing administration and a progressive climate of cross disciplinary research taking place on an international stage. For iSchools intent on educating leaders in the fields of records management, archival science and data curation it is imperative that these programs start to learn

from one another and look at themselves as being interdisciplinary rather than separate entities.

Cited Texts

[1] HEY, T., and A. E. TREFETHEN. 2003. The data deluge:An e-Science perspective. Grid Computing: Making the Global Infrastructure a Reality. Edited by F. Berman et al. ISBN 0470853190, pp. 809–824

[2] http://www.dcc.ac.uk/lifecycle-model/lifecycle_web.png

[3] http://www.dcc.ac.uk/resource/standards-watch/iso-15489/fig3.gif

[4] Schellenberg, T.R. "The Appraisal of Modern Public Records." In A Modern Archives Reader, edited by Maygene F.

Daniels and Timothy Walch, 57-70. Washington, DC: National Archives and Records Service, 1984.

[5] Harvey, Ross. (2007). Appraisal and Selection. Data Curation Manual, Digital Curation Centre.Available: http://www.dcc.ac.uk/resource/curationmanual/chapters/appraisal-andselection/appraisal-and-selection.pdf

The columns on the last page should be of equal length.

Organizing from the Middle Out: Citizen Science in the National Parks

Andrea Wiggins
Syracuse University
337 Hinds Hall
Syracuse, NY 13244
awiggins@syr.edu

ABSTRACT

This poster presents initial findings from a dissertation pilot study on a citizen science project involving the public with scientists in collaborative research. The goal for the pilot study was familiarity with the contextual factors that influence citizen science project design, and in turn, observing how the design choices contribute to the project's knowledge creation and participation outcomes. The initial results highlight an unexpected form of 'middle-out' organizing that challenges assumptions about top-down and bottom-up organizing, as the location of the top and bottom are clearly a matter of perspective in inter-organizational partnerships.

1. INTRODUCTION

This poster presents initial findings from a dissertation pilot study on a citizen science project involving the public with scientists in collaborative research. Many such projects are virtual organizations, with geographically dispersed resources and members who work toward common goals via cyberinfrastructure. Related research underscores the importance of understanding how organizational, task, and technology design requirements interact to influence participation and the scientific outcomes [5, 6].

Research which relies upon data about the natural world, and indeed the universe, is often hindered or rendered impossible by the high cost of data collection and analysis. The real-world problems that fall into this category depend on massive data sets that cannot be automatically generated, data collected over long periods of time or wide geographic areas, or large-scale analyses that require human perceptual competencies. The research problems range from climate change to the search for cures to cancer. To address these issues, as well as many other questions spanning a variety of disciplines, scientists are now employing citizen science as a solution to enable scientific research that is not feasible by any other means.

Public participation in scientific research can take a variety of forms; many of these projects resemble the Community Data and Open Community Contribution models of scientific collaboratories [2]. The dominant form of citizen science projects, found in the biological and environmental sciences, focuses on monitoring ecosystems and wildlife populations; volunteers form a human sensor network for distributed data collection [3, 1]. By contrast, in projects like NASA's Clickworkers [4], volunteers provide data analysis service, applying basic human perception to computationally difficult image recognition tasks.

Ubiquitous computing now makes broad public participation in scientific work a realistic research strategy for an increasing variety of projects. The evidence is clear that under the right circumstances, citizen science can work on a massive scale and is capable of producing high quality data as well as unexpected insights and innovations [1, 7], particularly when coupled with traditional scientific studies.

2. PILOT STUDY

The goal for the pilot study was familiarity with the contextual factors that influence citizen science project design, and in turn, observing how the design choices contribute to the project's knowledge creation and participation outcomes. The grand tour research question was, "How are citizen science projects formed?", and more specifically, "What factors influence the way a citizen science project develops?" The data collected between July and October of 2009 are currently under analysis, with initial findings reported here.

2.1 Study Site

The Northeast Phenology Monitoring (NPM) project is being developed by an inter-organizational network of partners collaborating virtually. They are working to create a regionally-coordinated citizen science project for implementation in National Parks in the Northeast region of the US. The goal of the project is to generate a large-scale phenological data set to study effects of climate change on natural life cycles in plants and animals. This pilot study focuses on the organizers who are designing the project, a virtual organization comprised of representatives from several organizations at multiple physical sites.

3. METHODS

The study employed ethnographic methods, taking an exploratory approach to develop a deeper understanding of the

context of a place-based citizen science project under development by a virtual organization. Data from interviews and participant observation are inductively coded with emergent themes relating to the processes and contextual factors influencing the development of the project. Data collection for the study began with seventeen interviews conducted by phone and in person; field notes generated detailed written observation, along with 315 digital images from the field sites and over 90 documents addressing many facets of the project's development.

4. FIELD SITES

Three field sites make up the locations for this pilot study. The first is virtual, and the other locations are physical sites in which the project's work is being implemented. The organizing group does most of its work virtually; coordination and communication occurs via email and phone, with periodic conference calls to report progress and plan next steps. The field sites for the citizen science project implementation were Acadia National Park and Boston Harbor Islands National Recreation Area.

Acadia was the first National Park east of the Mississippi, and today the Maine park encompasses 47,000 acres of unique mountain, forest, and seashore habitats. The park attracts several million visitors annually, offering 125 miles of hiking trails and 45 miles of carriage roads perfect for bicycling; it is widely considered a top birding destination in North America.

Boston Harbor Islands is a group of thirty-four islands in the Boston harbor, and was designated a National Recreation Area in 1996. The islands are owned and managed by a combination of private, public, state, and federal entities, operating together as the Boston Harbor Islands Partnership, and are home to several historic buildings, including a Civil War-era military fort and one of the nation's oldest lighthouses.

5. INITIAL FINDINGS

The project has no formal structure, and the common perception among members that this is not a 'top-down' project derives from the fact that there is no single dominant source of funding. Most organizational partners have little more invested than staff time, and as a result of low direct financial investment, the individual participants have greater autonomy to get the project's work done.

While the NPM is not a top-down project, neither is it 'bottom-up' in the usual sense of originating with the people who serve as volunteers. Instead, it seems to be an example of 'middle-out' organizing, in which the driving force in establishing the project partnership comes from the organizers who are positioned in between the funders and the volunteers, as shown in Figure 1.

The lack of centralized funding creates conditions that permit organizers to contribute as they are best able while maintaining a focused scientific goal, which may be less likely to occur in bottom-up scenarios. Although the organizers who are employed by federal agencies view their effort as ground-up, they overlook volunteers as potential instigators, which is why I propose that this project demonstrates middle-out organizing:

> "The lack of funding from the top, that isn't going to come, so they're [regional managers] going to have to be the creators, they're going to have to be their own little generators to get it up and running. And it'll become institutionalized because a number of people will just make it happen, and that's how, in a sense, it's a ground-up instead of top-down sort of a thing."

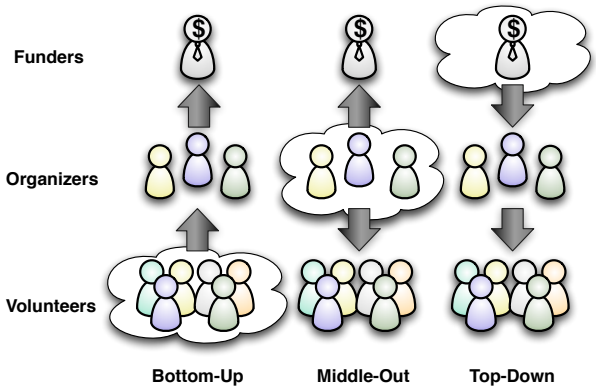


Figure 1: The organizational location of the project's originators shape its organizing processes.

This project emerged in a fiscal environment that imposed severe limitations, but the group members have found that pulling together just enough funding from a variety of sources has permitted a different approach than they would otherwise expect. As the de facto project leader observed:

> "So what you have here is a project that's being done loosely among a bunch of different organizations. Typically in a project like this, you'd have a primary source of funding kind of providing structure to it all. But we don't have that here at all. So it's just kind of everybody participating as they can, and as they have time, and as they will."

Of course, the lack of funding is never considered a good thing. It provides some advantages to the organizers in terms of their autonomy, which comes at the steep cost of personal commitment. According to a park staff member,

> "Mostly it's the individual biologist at that park who's willing to say, my forty hours a week is here, but I'm going to spend my extra eight hours a week, and this is going to be my baby because I feel something strong, and that's how it goes...So there's no national coordinator, there's no anything. But there again, I think we can be our own independent people and do our own thing."

Engaging in a middle-out partnership like the NPM project is a matter of balance, but the organizers are enthusiastic and committed despite many constraints. They see this

project as fulfilling a key organizational goal with respect to managing natural resources, and also addressing a particularly pressing environmental issue by generating the data needed for the National Park Service to make scientifically-based managerial decisions in response to climate change.

6. CONCLUSIONS

The initial analysis of the NPM project touches on challenges and advantages of organizing a citizen science project partnership from the middle outward. The NPM project has emerged from a small group of committed individuals across a variety of organizational locations scattered somewhere in between the top and bottom. These findings challenge our assumptions about top-down and bottom-up organizing, as the location of the top and bottom are clearly a matter of perspective in inter-organizational partnerships.

7. ACKNOWLEDGMENTS

This work is partially supported by NSF Award # 0943049 under the American Recovery and Reinvestment Act of 2009.

8. REFERENCES

[1] R. Bonney and M. LaBranche. Citizen science: Involving the public in research. *ASTC Dimensions*, page 13, May/June 2004.

[2] N. Bos, A. Zimmerman, J. Olson, J. Yew, J. Yerkie, E. Dahl, and G. Olson. From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication*, 12(2):652–672, 2007.

[3] J. P. Cohn. Citizen science: Can volunteers do real research? *BioScience*, 58(3):192–107, March 2008 2008.

[4] B. Kanefsky, N. Barlow, and V. Gulick. Can Distributed Volunteers Accomplish Massive Data Analysis Tasks? *Lunar and Planetary Science*, 1, 2001.

[5] C. Lee, P. Dourish, and G. Mark. The human infrastructure of cyberinfrastructure. In *Proc. CSCW 2006*, pages 483–492, 2006.

[6] S. Star and K. Ruhleder. Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. In *Proc. CSCW 1994*, pages 253–264, 1994.

[7] D. Trumbull, R. Bonney, D. Bascom, and A. Cabral. Thinking scientifically during participation in a citizen-science project. *Science Education*, 84(2):265–275, 2000.

Sensemaking in the Space: An Alternative Design Perspective for Mobile Navigation Systems

Anna Wu
Pennsylvania State University
IST Building 327
University Park, PA
1-814-863-6822
annawu@psu.edu

Xiaolong(Luke) Zhang
Pennsylvania State University
IST Building 307D
University Park, PA
1-814-863-9462
lzhang@ist.psu.edu

John M. Carroll
Pennsylvania State University
IST Building 307H
University Park, PA, USA
1-814-863-2476
jcarroll@ist.psu.edu

Alexander Klippel
Pennsylvania State University
Walker Building 204
University Park, PA, USA
1-814-865-2324
klippel@psu.edu

ABSTRACT

In this paper, we present the essential idea of an on-going project studying navigation in physical environment as sensemaking process. Initial design guidelines are proposed for discussion.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces – Evaluation/methodology, Prototyping.

General Terms

Theory, Design

Keywords

Navigation, sensemaking, mobile interface design

1. Poster Abstract

Navigation in physical environments is a basic human survival skill. We study maps before set outting and connect 3D real world objects with 2D abstract icons on maps during the navigation by continuously checking and consciously memorizing. Complex modern architectures and city plans plus increasing travel span challenge our navigation ability. If finding our way is solving a spatial problem as Passini suggested more than 30 years ago [1] and has been believed ever since, nowadays, this problem becomes more and more difficult.

Proliferation of GPS units simplifies navigation in unfamiliar places as following step-by-step direction. We no longer need to go through the whole process of spatial information gathering, decision making, and decision executing as problem-solving. The active explorer is degenerated into a passive follower. Problem coming with such cognitive easiness is the "mindless of the environment"[2]. In case GPS devices are out of access, malfunction, or simply give wrong directions, people may not be well prepared to react to unexpected environmental conditions and find alternative action plans. Empirical study suggested that with the slavery of automatic tools could results in degeneration in acquisition of spatial knowledge [2].

While most current studies on navigation come from the perspective of goal-reaching, but the contribution of navigation to spatial knowledge learning is under-investigated. The process can enhance learning and knowledge gaining even the problem is not solved, or not quickly solve. Sensemaking, a concept first proposed in 1980s [3], was reproposed and has become a serious study field [4-7] triggered by information explosion: we need to find meaning of the world regardless the increasing data volume. Several models are proposed to capture sensemaking process in both individual level and organizational level[e.g. 4, 5]. Though vary in details, most of these models agree that sensemaking is an iteratively engaging process that tries to bridge the gap between observed information with structured concepts (e.g. encoding data with schema, instantiating structure) in order to form a coherent understanding. In such iterations, computational tools that provide proper external representations are believed to facilitate individual sensemaking process by reducing transaction memory, influencing the level of participation, providing manageable artifacts, and helping pattern recognition, which is highly desired at current stage[8, 9].

We propose another perspective of physical navigation as a sensemaking process. Despite physical navigation is a direct metaphor for making sense out of massive information, fewer researchers approach navigation from this perspective except: Kevin Lynch [10] introduced a viewpoint to how inhabitants interpret environment with his pioneering work on imageability, which characterizes the way people create mental pictures consisting spatial primitives (paths, edges, districts, nodes, and landmarks); Klein argued that the lost and recovery stage in navigation could be treated as sensemaking processes based on his semi-formal interview [11]. Moreover, no practical design supports navigation as to support spatial sensemaking, considering the scarce existing tools support sensemaking in general.

Viewing navigation as a legitimated sensemaking in scenarios where cognitive agents need to know the environment instead of a one-time visit, we are investigating the design implications for mobile guides. Particularly, how to support sensemaking during navigation by visual representations with the limited display estate and cognitive attention during moving? What kind of information, in what way to present? "How much is too much?" When comes to artifact design as mobile navigation guide, the problem is how to balance cognitive cost with spatial awareness of the physical environment [e.g. 2, 12].

Based on previous work in both sensemaking and physical navigation, we present a sensemaking framework and the analysis of navigation as spatial sensemaking process with respect to spatial information and clue collection, options and choices comparison at decision making points, fragmented and cognitive maps formation. These processes are similar to the iternarative process of fit data into frame and forming frame from data loop in sensemaking. In physical navigation, people comprehend and structure their experiences into imaginative mental representations as they moving and interacting with the environment. In order to effectively identify relevant cues in the environment, navigators collect and filter massive spatial information based on objects' saliency, either perceptively or cognitively, which reflects the visual prominence or recognized importance for forming mental representation. These salient objects, as anchors, work along with paths connecting relevant places to form fragmented representations. Integrating these fragmentations into a coherent cognitive map by identifying joint points and correct orientation is difficult process. Once the imagery (or image schemata[13]) is formed, such recurring mental patterns could help people learn the space and make qualitative judgment and essential reasoning.

Finally, the following design implications are proposed for interface design on mobile navigation systems.

+ Visualizing landmarks
  - Provide perceptive and cognitive salient objects as anchor points to pin mental structure.
  - Provide salient entities (e.g., important landmarks) to connect the anchor points in different scales.
+ Visualizing routes
  - Provide generalized route maps which emphasizes on segmentation and branches, rather than distance.
  - Provide aspect route maps for different purposes
  - Provide feedback after decision made at decision point for confirmation or instant reorientation.
+ Visualizing surrounding context
  - Provide a topological relationship view that preserve environmental patterns (e.g., the spatial relationship among cognitively salient entities and routes) to help construct the frames of mental image.
  - Provide different "level of detail" maps results in mental models with different completeness to response to different task requirements.
  - Provide content-context visualization to support connecting fragmentation into complete mental image. (coordination transformation)

2. REFERENCES

[1] Passini, R. Wayfinding: A study of spatial problem solving with implications for physical design. Pennsylvania State University, State College, 1977.

[2] Parush, A., Ahuvia, S. and Erev, I. Degradation in Spatial Knowledge Acquisition When Using Automatic Navigation Systems. In S. Winter (eds). LNCS. Heidelberg: Springer Berlin, 2007.

[3] Dervin, B. An overview of Sense-Making research: Concepts, methods, and results to date. the annual meeting of the International Communication Association (Dallas, TX, 1983).

[4] Weick, K. E. Sensemaking in organizations. Sage Publications, 1995.

[5] Russell, D. M., Stefik, M. J., Pirolli, P. and Card, S. K. The cost structure of sensemaking. Proc. INTERCHI'93, 1993. 269-276.

[6] Klein, G., Phillips, J. K., Rall, E. L. and Peluso, D. A. A Data-frame theory of sensemaking. Proc. 6th International conference on naturalistic decision making, 2006.

[7] Pirolli, P. and Card, S. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Proc. International Conference on Intelligence Analysis, 2005.

[8] Klein, G., Moon, B. and Hoffman, R. R. (2006) Making Sense of Sensemaking 2: A Macrocognitive Model. IEEE INTELLIGENT SYSTEMS, 21(5), 88-92.

[9] Faisal, S., Attfield, S. and Blandford, A. A Classification of Sensemaking Representations. CHI'09 Sensemaking Workshop. (Boston, MA, USA, 2009).

[10] Lynch, K. The Image of the City. MIT Press, Cambridge, MA, 1960.

[11] Klein, G. Corruption and recovery of sensemaking during navigation. In M. J. Cook, J. M. Noyes and Y. Masakowski (eds). Decision making in complex environments. 13-32. Burlington, VT, USA: Ashgate, 2007.

[12] Forbes, N. Online survey of in-vehicle navigation system users. University of Nottingham. , City, 2006.

[13] Johnson, M. The body in the mind: The bodily basis of meaning, imagination, and reason. University of Chicago Press Chicago, 1987.

Embodying Research Methods into Fields and Tables: A Process of Informed Database Design

L. Wynholds
wynholds@ucla.edu

David Fearon
dfearon@ucla.edu

Christine L. Borgman
borgman@gseis.ucla.edu

Sharon Traweek
traweek@history.ucla.edu

Department of Information Studies
University of California, Los Angeles

ABSTRACT

One of the invisible aspects of large research projects in the social sciences is the method by which observations and other collected data are managed. In sufficiently large projects, it may be effective to address the data management problem at the outset by creating a database architecture and data processing workflow. Research methods, assumptions and technical limitations often drive the structure of the data to be collected, but this is rarely discussed within the framework of the research. This design process represents a complex selection and trade-off matrix of predictive approximation, given that aspects of the analysis are not performed until the data is collected, and the design is done before the data collection is started. An elegant design can afford an equally elegant analysis of the data, but also creates a cycle where the data structure dictates the focus and granularity of the analysis.

We were faced with the problem of creating a system to support the projected data collection projects for a major, multi-method, 5-year research project on data curation practices. Our research focuses on specific techno-social practices of astronomers and will rely on a large volume of complex and heterogeneous source materials, such as email archives, scholarly publications, websites, reports, metadata headers, as well as in-person interviews.

The research questions focus on the data management, curation, and sharing practices of astronomers, how these practices evolved, and mapping who shares what, when, with whom, and why, with specific interest in what data they generate, use, keep and discard. We also ask what is most important to curate, and how do they do so, what do they expect to use and decide will be of future use to others, and who do they envision as future users?

The database structure will act as the connective tissue for the full term of the project while embodying the research methods, facilitating analysis, enabling data sharing, and minimizing effort. However, the process also represents a complex selection and trade-off matrix of predictive approximation of the intended analysis as the design defines the data set and the data set drives the analysis. This process-oriented poster documents the matrix we followed, the challenges and the solutions developed while operationalizing a data system for a large research project with major relational and descriptive aspects. Our resulting system utilizes existing competencies and departmental resources while meeting basic prerequisites for data security, sharing, interoperability, best practices and extensibility.

Automatic Extraction of Location Relations from Text

Wu Zheng
University of Illinois at Urbana Champaign
wuzheng2@illinois.edu

Catherine Blake
University of Illinois at Urbana Champaign
clblake@illinois.edu

ABSTRACT

Automatically identifying semantic relationships from text plays an important role in knowledge discovery, for example to connect a researcher in one discipline to related research questions in a second discipline in which the researcher is not formally trained. This poster describes preliminary experiments in ongoing research project that explores the utility of semantics and syntax to identify relations from text automatically. We focus exclusively on the *location* relation, such as organization-location and gene-location. Location is an interesting case because it occurs in multiple text genres including news articles and scientific literature.

Keywords

Information extraction, relation mining, text mining, knowledge discovery

INTRODUCTION

The extraction of semantic relations plays an important role in knowledge discovery from text. Identifying relationships automatically can contribute to a variety of natural language processing applications, including information extraction, question answering, knowledge discovery, and information synthesis (Lapata, 2002; Morris & Hirst, 2004; Rosario & Hearst, 2005). The research presented in this poster is part of a broader project entitled Evidence-Based Discovery (EBD) where the goal is to enable a scientist in one domain to more easily identify findings from another domain in which the scientist has little or no formal training. Our goal in the broader project is to develop language technologies that automatically identify a range of semantic relations from text.

In this poster, we present preliminary results of experiments that explore the utility of semantics and syntactic constraints to identify a location relationship. The location relation is an interesting case because it spans multiple genre's for example identifying the geographical location of an organization's head office, the location of a gene within the body, or the location of a city within a country. Smith et al (Smith, et al., 2005) characterized location as a primitive instance-level relation in their research on relations in biomedical ontologies.

We define the location relation as a binary predicate where the arguments define the item of interest (in the previous examples the organization, the gene or the city) and the location of that item, specifically LOCATION(X, Y) indicates that X is located in Y. In sentence (1) below the system should instantiate the location relation as LOCATION(Slr0228, thylakoid membrane).

> Given the predicted orientation of these helices and assuming that Slr0228 is located in the thylakoid membrane, the conserved 81-amino acid feature would constitute a lumenal domain. (1)

METHOD

One of the fundamental tenants of computational linguistics is that there exists a relationship between the underlying form (the syntax) and the meaning conveyed (the semantics) in a sentence. Our goal is to explore the degree to which syntactic and semantic features of a sentence can enable us to identify new terms that are indicative of a location relation. We use a similar strategy to the development of the FrameNet database (see (Gildea & Jurafsky, 2001), pg 5).

(1) Identify seed terms for the binary predicate

(2) Sample sentences with the seed terms and identify arguments

(3) Characterize syntactic constructs

(4) Check the annotated sentences for consistency

(5) Use arguments to identify new seed terms for the binary predicate and repeat steps 1-4.

The first semantic constraint was the seed word "located" followed by the prepositions "in", "at", and "on". We drew a stratified random sample of 100 sentences that contained each of the three seed phrases from the 1.8 million sentences in the 2002 Journal of Biological Chemistry (JBC) articles from TREC (Hersh & Voorhees, 2009). Sentences in each category - located in, located at, and located on - were sampled at a rate proportional to their overall distribution in the journal (see Table 1).

	Sample	Journal
located in	58	3,528
located at	26	1,665
located on	16	834
Total	100	6,027

Table 1: Number of sample sentences in each category

With the semantic constraints in place, the next step is to identify syntactic patterns that characterize the location relations. The syntactic patterns explored in this paper are the typed dependencies produced by the Stanford Parser (Klein & Manning, 2003). Of the 1.8 million sentences in JBC, a parse was produced for 1.66 million (92%).

For step 3, we manually inspected each sample sentence to identify syntactic patterns that are indicative of location relations. Figure 1 shows the dependency tree for sentence 1, where location (X,Y) should return X=Slr0228 and Y=thylakoid membrane. In this case, the passive nominal subject (nsubjpass) identified by the Stanford parser corresponds directly to first argument of the location relation and the object of the preposition (pobj) corresponds directly to the second argument of the location relation. Even though the subject and object returned by the Stanford parser do not necessarily reflect the subject and object of location verb we refer to X as the subject and Y as the object of the location relation to ease further discussion.

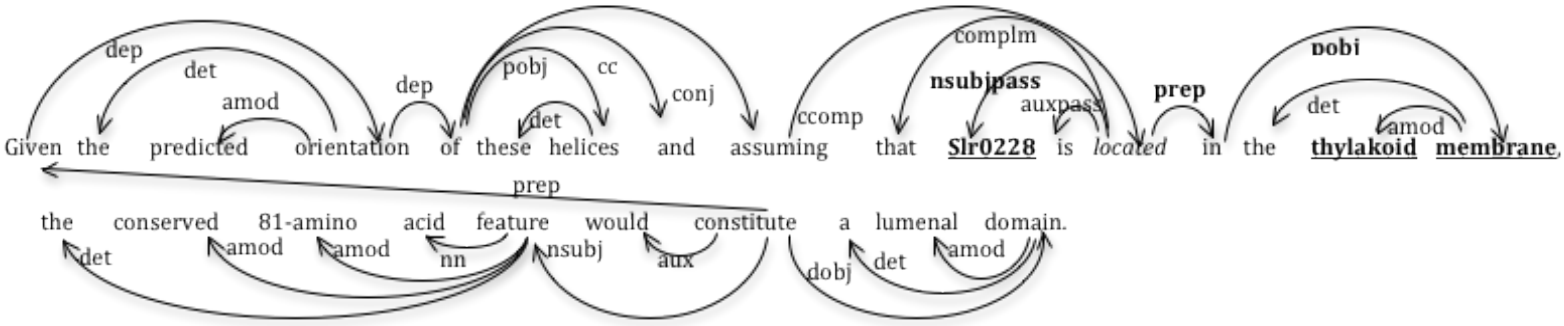


Figure 1: Dependency tree for sentence (1)

RESULTS AND DISCUSSION

Our analysis of the 100 sample sentences revealed two syntactic patterns that are highly indicative of location relations.

(1) Terms depicted as subjects of the predicate seed term. For example, in the following sentence, the phrase *the catalytic serine (Ser_153)* corresponds to the first argument in location.

> The catalytic serine (Ser_153) is in a characteristic epsilon conformation and is located in a tight turn with the G-H-S-Q-G sequence belonging to the usual consensus sequence of the alpha/beta hydrolase fold family. (2)

(2) The seed term is a participial modifier of the first location argument. For example in the sentence (3), *located* is the participial modifier of term *AU-rich elements (ARE)*.

> Decay of these cytokine mRNAs is normally regulated in part by the presence of AU-rich elements (ARE) located in their 3'-untranslated regions. (3)

We also observed variations of these two rules. For example, in some sentences the seed term was connected to the location arguments via a conjunction. In these cases, the Stanford parser associated the subjects of the sentence with just the first verb and the subject of our seed term was not represented directly in the dependency tree. In these sentences, the first location argument can be identified by tracing back though the dependency tree to identify the subject. For example in sentence (4) *Slt2-GFP* is the first located argument. In the output of the parser, *Slt2-GFP* is the subject of the verb *concentrated*, but the dependency tree connects the second argument (cytoplasm) via the conjunction *and*.

> In accordance with the previous report localizing Slt2-HA, Slt2-GFP was concentrated in the nucleus and also located in the cytoplasm at 25 C. (4)

Another variation is that the seed term appeared in the open clausal complements of other verbs in a sentence. (An open clausal complement is a clausal complement without its own subject, whose reference is determined by an external subject.) In this case, the first argument of the location relation is the subject of the verb that the open clausal complement modifies. For example in sentence (5), the dependency parser comprises list the phrase *some of the ER retention signals* as the subject of the verb *shown* and the seed term *located* is the open clausal complement of *shown*.

> Since some of the ER retention signals have been shown to be located in the C-terminal end of polypeptides, we generated constructs with deletion of 15, 40, and 97 amino acids from the C-terminal tail of RyR. (5)

In contrast to the syntactic patterns that would accurately identify the first argument of the location relation, we found only one syntactic pattern that worked surprisingly well for the second argument. Specifically, the head noun identified by the Stanford parser for the seed term was typically the second argument in the located relation. For example (1), the phrase *the thylakoid membrane,* which is the object of *located in*.

We identified rules that would identify both the predicate and arguments for the location relation. Errors from seven of the 100 sentences were caused by incorrect parsing. The performance on the development collection of sentences is shown in Table 2. The rules were evaluated using two different criteria. Under the strict criterion, we considered an extracted term correct if and only if the term matched the whole phrase of each of the location relation arguments. Under the loose criterion, we considered an extracted term correct if it captured the head noun of location relation arguments. In the later case missing modifiers were not considered errors.

The evaluation of the development collection suggests that syntactic features are highly effective in capturing arguments of the location relation. We are currently working on steps 4 and 5, which will establish the consistency of these relations and use the arguments identified from the seed term located, to generate additional verbs that are indicative of location relation.

			Correct Counts		Precision		Recall	
	Number of Prediction	Actual Number	Strict	Loose	Strict	Loose	Strict	Loose
First argument	101	103	66	89	65.3%	88.1%	64.1%	86.4%
Second argument	122	103	71	100	58.2%	82.0%	68.9%	97.1%
Both arguments	223	206	137	189	61.4%	84.8%	66.5%	91.7%

Table 2: Performance of rules on development collection

Our approach does depend on the performance of the parser. Sentences from scientific literature, which are quite different from the sentences in newspaper articles on which the Stanford parser was trained, in particular they tend to be longer and more complex. Training the parser on scientific literature may further improve the system results.

ACKNOWLEDGEMENT

This material is based in part upon work supported by the National Science Foundation under Grant IIS-0812522. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Gildea, D., & Jurafsky, D. (2001). Automatic labeling of semantic roles. *Computational Linguistics, 99*(9), 1-43.

Hersh, W. R., & Voorhees, E. M. (2009). TREC genomics special issue overview. *Information Retrieval, 12*(1), 1-15.

Klein, D., & Manning, C. D. (2003). *Fast Exact Inference with a Factored Model for Natural Language Parsing.* . Paper presented at the Advances in Neural Information Processing Systems.

Lapata, M. (2002). The disambiguation of nominalisations. *Computational Linguistics, 28*(3), 357–388.

Morris, J., & Hirst, G. (2004). *Non-classical lexical semantic relations*. Paper presented at the Proceedings of the 4th Human Language Technology Conference and 5th Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2004) - Workshop on Computational Lexical Semantics.

Rosario, B., & Hearst, M. (2005). *Multi-way relation classification: application to protein-protein interaction.* Paper presented at the Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing Vancouver, British Columbia, Canada

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., et al. (2005). Relations in biomedical ontologies. *Genome Biology, 6*(5), R46.