

iSCHOOLS CAUCUS PRESENTS THE THIRD ANNUAL

iConference 2008

iFutures: Systems, Selves, Society

February 28 - March 1, 2008

UCLA :: Los Angeles, CA

Welcome to iConference 2008!

Please open the “Bookmarks” tab on the left-hand side of this screen, and use that menu to navigate through this document.

This is your conference schedule, guide, and reference tool.

Here you'll find:

- The full submissions and abstracts of every poster, paper, wildcard, and roundtable session at the conference.
- Daily conference schedules with descriptions of all the activities and links to the presenter's submissions.
- A guide to UCLA and the surrounding area, including maps, local attractions, UCLA events, and a restaurant list.

Abbreviated Schedule

Wednesday, February 27th

Pre-conference Opportunities

1 – 5 pm Doctoral Colloquium • Session Chair: Jonathan Furner
The Colloquium will bring together a selected group of outstanding doctoral students (with *i* School-type interests but not necessarily in a school that is a member of the *i* School Caucus) for workshops with faculty.

- Covell Commons, West Coast Room & South Bay Room

5:15 – 8 pm Doctoral Colloquium Dinner/Evening Program

Session Chair: Jonathan Furner

Dinner for colloquium members followed by additional activities.

- Covell Commons, West Coast Room & South Bay Room
-

Thursday, February 28th

iSchools Conference 2008

8 am – noon Conference Registration

Participants will pick up their conference packets, final agenda, and name badges just outside of the Grand Horizon Room on the 3rd floor of Covell Commons.

- Covell Commons, Grand Horizon Room

9 – 11:30 am Poster Sessions

Session Chair: Jonathan Furner

Authors will present 30 posters related to new and promising research results and/or current design projects. Posters will remain available for viewing from 11:30 am to 5:30 pm. Posters must be taken down between 5:30-5:45 pm on Thursday, February 28.

- Covell Commons, Northridge Room

**noon - 1:30 pm Conference Welcome Luncheon
with Keynote Speaker Elizabeth Liddy**

UCLA Graduate School of Education & Information Studies Dean

Aimée Dorr will welcome attendees to this year's iSchool conference and introduce the keynote speaker, Elizabeth Liddy, Dean of the School of Information Studies at Syracuse University.

- Covell Commons, Grand Horizon Room

1:30 - 2 pm Afternoon Break

- Covell Commons, Grand Horizon Terrace

2 – 3:30 pm Early Afternoon Breakout Sessions

Topics will range from information technologies and networking to information law and preservation. We expect more than 20 individual presentations exploring the ten conference themes and more.

- Paper Session #1: Session Chair: John Unsworth
- Paper Session #2: Session Chair: Jenny Preece
- Wildcard Session #1: Jimmy Lin
- Wildcard Session #2: Jean-François Blanchette
- Wildcard Session #3: Alpha Selene DeLap
- Wildcard Session #4: David Ribes

- Northridge Room (30- Posters Remain on Display)

3:30 – 4 pm Afternoon Break

- Covell Commons, Grand Horizon Terrace

4 – 5:30 pm Junior Faculty Mentoring

Session Chairs: Leah A. Lievrouw & Steve Sawyer

By invitation only.

- Bradley Room 213/214

4 – 5:30 pm Late Afternoon Breakout Sessions

Sessions continue to explore the conference themes and beyond. We also plan to include a junior faculty mentoring activity during this time.

- Paper Session #3: Session Chair: Brian Cantwell Smith
- Wildcard Session #5: Ramesh Srinivasan
- Wildcard Session #6: Derek Hansen
- Wildcard Session #7: Tatjana Aparac-Jelusic
- Wildcard Session #8: Richard James Cox
- Northridge Room (30- Posters Remain on Display)

5:30 – 7 pm Microsoft® Reception

Join colleagues, mentors, and researchers on the Grand Horizon Terrace overlooking the UCLA campus. Refreshments and hors d'oeuvres will be served.

- Covell Commons, Grand Horizon Terrace

7 pm Dinner on your own

UCLA buses will transport you to your hotel or to one of more than a dozen restaurants in Westwood. Or make an optional visit to UCLA's Hammer Museum with an outstanding collection of both contemporary and traditional art. The return trip to your hotel from a restaurant or the Hammer will be on your own.

Friday, February 29th

iSchools Conference 2008

8 - 8:30 am Continental Breakfast

- Covell Commons, Grand Horizon Terrace

8:30 - 10 am Early Morning Concurrent Sessions

The early morning concurrent sessions continue to explore the conference themes and beyond.

- Paper Session #4: Session Chair: Marcia Bates
- Paper Session #5: Session Chair: Olivia Frost
- Paper Session #6: Session Chair: Stephen Hirtle
- Roundtable Session #1: Terry Weech
- Roundtable Session #2: Clara M. Chu
- Roundtable Session #3: Andrew T. Fiore
- Roundtable Session #4: Scott Paquette
- Roundtable Session #5: Deborah Turner
- Wildcard Session #9: Alla Zollers
- Wildcard Session #10: Susan Leigh Star

10 - 10:30 am Morning Break

- Covell Commons, Grand Horizon Terrace

10:30 - Noon Late Morning Concurrent Sessions

The late morning concurrent sessions continue to explore the conference themes and beyond.

- Paper Session #7: Session Chair: Harry Bruce
- Roundtable Session #6: Scott Nicholson
- Roundtable Session #7: Eric Christopher Kansa
- Roundtable Session #8: Carol L. Tilley
- Roundtable Session #9: Melissa Rita Ho
- Roundtable Session #10: Charles Carlton DuBois
- Wildcard Session #11: Jim Jansen
- Wildcard Session #12: M. Cameron Jones
- Wildcard Session #13: Michael D. Crandall
- Wildcard Session #14: Daniela Karin Rosner

Noon - 1:30 pm Lunch

- SV Dining Hall

&

Junior Faculty Mentoring and Lunch

Session Chairs: Leah A. Lievrouw & Steve Sawyer

By invitation only.

- De Neve Plaza A

&

Doctoral Colloquium Group Lunch

By invitation only.

- Sunset Village Dining Hall

1:30 - 3 pm Early Afternoon Concurrent Sessions

The early afternoon concurrent sessions continue to explore the conference themes and more.

- Paper Session #8: Session Chair: Debra Richardson
- Paper Session #9: Session Chair: Bruce Kingma
- Paper Session #10: Session Chair: Doug Oard
- Roundtable Session #11: Nancy Van House
- Roundtable Session #12: Saira Naim Haque
- Roundtable Session #13: John Yen
- Roundtable Session #14: Sean A Munson
- Wildcard Session #15: Frederic D Stutzman
- Wildcard Session #16: Indira Rita Guzman
- Wildcard Session #17: Michael B. Eisenberg

3 pm The Getty Center

Board buses for a visit to the Getty Research Institute and an evening at the Getty Museum. This special event features a program from Getty Research Institute speakers, followed by a reception. Conference participants may visit the GRI and Museum galleries until 9 pm.

RSVP is required as there is limited seating.

Please email Angela DeCenzo ASAP to reserve your space.

Saturday, March 1st

iSchools Conference 2008

- 8 - 9 am Continental Breakfast**
• Northwest Campus Auditorium Patio
- 9 - 10:30 am Keynote Presentation by**
Christa Phillips, Xbox.com, Microsoft® Corporation
• Northwest Campus Auditorium
- 10:30 - 11 am Morning Break**
• Northwest Campus Auditorium Patio
- 10:30 - 11 am Plenary Concluding Sessions**
• Northwest Campus Auditorium

ABOUT *i*SCHOOLS

The *i* Schools Caucus consists of schools interested in the relationship between information, technology, and people. This is characterized by a commitment to learning and understanding of the role of information in human endeavors. The *i* Schools Caucus works toward this vision that emphasizes connecting people with information by designing, developing, and evaluating information technologies, systems and services and they bring together faculty from a range of more traditional disciplines to conduct multidisciplinary research. The *i* Schools take it as a given 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, as well as information technologies and their applications.

i Conference Planning Committee:

- Aimée Dorr, University of California, Los Angeles
- Tom Finholt, University of Michigan
- Anne Gilliland, University of California, Los Angeles
- Jose-Marie Griffith, University of North Carolina, Chapel Hill
- Bonni Nardi, University of California, Irvine
- Allen Renear, University of Illinois, Urbana-Champaign
- Debra Richardson, University of California, Irvine
- Steve Sawyer, Pennsylvania State University
- Paul Solomon, University of North Carolina, Chapel Hill
- John Unsworth, University of Illinois, Urbana-Champaign

The *i* Schools Caucus:

University of California, Berkeley - School of Information

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

Drexel University - College of Information Science and Technology

Florida State University - College of Information

Georgia Institute of Technology - College of Computing

University of Illinois Urbana-Champaign - The Graduate School of Library and Information Science

Indiana University - School of Informatics

Indiana University - School of Library and Information Science

University of Maryland - College of Information Studies

University of Michigan - The School of Information

University of North Carolina - School of Information and Library Science

The Pennsylvania State University - College of Information Sciences and Technology

University of Pittsburgh - School of Information Sciences

Rutgers, the State University of New Jersey - School of Communication, Information, and Library Studies

Syracuse University - School of Information Studies

University of Texas, Austin - School of Information

University of Toronto - Faculty of Information Studies

University of Washington - Information School

If you are interested in joining the *i* Schools Caucus, please contact Harry Bruce at harryb@u.washington.edu

Giving your work more visibility

Through the online **eScholarship Repository** and the open-access, peer-reviewed journal **InterActions**, your work can be found more easily by researchers. Your accepted paper will automatically be deposited in the repository and posted to the iConference Web site, and you are encouraged to submit it to **InterActions** as well.



Collaboratively edited by UCLA graduate students in both Education and Information Studies, **InterActions** is a peer-reviewed on-line journal of critical scholarship committed to multidisciplinary perspectives and interdisciplinary communication.

Submissions to the special iConference2008 issue of **InterActions** may be made at our Web site at <http://www.interactions.gseis.ucla.edu>. Please direct any questions regarding the special issue to Stacey Meeker (smeecker@ucla.edu). Publications can include a range of media and data files.



With your accepted submission, your paper will be deposited and preserved in the **California Digital Library's eScholarship Repository**, an open-access publishing platform for scholarly output.

If you would like to add other materials to supplement your paper such as handouts or audio/visual files, please email the Digital Collections Services Librarian at the **UCLA Library**, Bonnie Tijerina (btijerina@library.ucla.edu).

More information about the **eScholarship Repository** is available at: <http://repositories.cdlib.org>.



Welcome to the 2008 Third Annual **i**Schools Conference, **iFutures: Systems, Selves, Society**, bringing together scholars and professionals who are working at the nexus of information, technology, and people.

The **i**Schools Caucus and the **i**Conference Planning Committee are delighted to offer more than 90 presentations, including invited speakers, paper sessions, posters, roundtables, and wildcards. A doctoral colloquium, mentoring for junior faculty, and a special visit to the Getty Research Institute will provide additional opportunities for learning and sharing, celebrating our multidisciplinary efforts to understand and create the future roles of information in our society.

The UCLA Graduate School of Education & Information Studies is honored to host this year's event. May your time here be stimulating and rewarding.

AIMÉE DORR

Dean, UCLA Graduate School of Education & Information Studies

keynote speakers



ELIZABETH D. LIDDY

Featured Guest Speaker, Thursday, February 28th

Elizabeth D. Liddy is Interim Dean and Trustee Professor in the School of Information Studies at Syracuse University. In 1999, she was appointed director of the school's Center for Natural Language Processing, which advances the development of human-like language understanding software capabilities for government, commercial and consumer applications. Liddy is also an adjunct professor at Upstate Medical University, where she conducts research on medical informatics. Liddy has led 65 research projects, with the support of numerous government agencies and commercial enterprises. In addition, she is a co-inventor on five patents in the area of natural language processing (NLP). Among her many honors, Liddy is a recipient of the Tibbetts Award from the SBIR Program of the U.S. Small Business Administration (1998), the Enterprise Award for Technology from the Upstate New York Technology Business Founm (1998), and the Outstanding Alumni Award from Syracuse University (2000). She holds a bachelor's degree in English Language and Literature from Daemen College and an M.L.S. in Information Studies and a Ph.D. in Information Transfer, both from Syracuse University.



CHRISTA PHILLIPS

Featured Guest Speaker, Saturday, March 1st

Christa Phillips, better known by her online persona "TriXie," is the Community Editor at Xbox.com where she manages columnists, interviews gamers, covers industry events such as E3 and Tokyo Game Show, and runs Xbox LIVE programming events Frag Doll Friday, Family Game Night and Xbox All-Nighter. She also manages and co-founded GamerchiX, an international group for women who play video games. Since the group began in June 2006, GamerchiX membership has grown to over 4,300 women in 32 countries. GamerchiX has been featured in the *Los Angeles Times*, *The Escapist.com*, and *The Chicago Tribune*. Phillips has been in the games industry for 13 years. She began her career as an assistant editor at *InterAction* magazine, and then became a packaging and advertising copywriter at Sierra On-Line before coming to Xbox in 2002 where she worked on the launches of Xbox LIVE, Xbox 360, Halo 2 and Halo 3. Her interests include reading, music, raising her two kids, and of course, playing video games.

acknowledgements

This year's conference was made possible only through the dedicated work and support of many individuals and organizations. A very special thanks to every individual who submitted a proposal to the conference and to our dedicated proposal reviewers. Thanks as well to the following:

SPONSORS



National Science Foundation

Microsoft®

GETTY MUSEUM PRESENTATION AND RECEPTION

Susan Allen
Kenneth Hama

iSCHOOLS CONFERENCE PLANNING COMMITTEE, SPECIAL THANKS:

Anne Gilliland, University of California, Los Angeles
Debra Richardson, University of California, Irvine
Paul Solomon, University of North Carolina, Chapel Hill
John Unsworth, University of Illinois, Urbana-Champaign

DOCTORAL COLLOQUIUM

Jonathan Furner, University of California, Los Angeles

JUNIOR FACULTY MENTORING

Leah Lievrouw, University of California, Los Angeles
Steve Sawyer, Pennsylvania State University

CONFERENCE PLANNING AND GRAPHICS

Ron Dietel, University of California, Los Angeles
Angela De Cenzo, University of California, Los Angeles
Judy K. Lee, University of California, Los Angeles
Leslie Lewis, University of California, Los Angeles

table of contents

Welcome :: pg 2

Keynote Speakers :: pg 2

Acknowledgements :: pg 3

Wed. Schedule :: pg 3

Thurs. Schedule :: pgs 4-5

Fri. Schedule :: pgs 5-7

Sat. Schedule :: pg 7

Details on Presentations

Posters :: pgs 8-9

Papers :: pgs 10-11

Wildcards :: pgs 12-13

Roundtables :: pgs 13-14

Maps :: pg 15, Back Cover

Wednesday, February 27th

PRE-CONFERENCE OPPORTUNITIES BY INVITATION

1:00 – 5:00 pm **Doctoral Colloquium supported by National Science Foundation**

Session Chair: Jonathan Furner

The Colloquium will bring together a selected group of outstanding doctoral students (with *i*School-type interests but not necessarily in a school that is a member of the *i*School Caucus) for workshops with faculty.

- Covell Commons, West Coast Room & South Bay Room

5:15 – 8:00 pm **Doctoral Colloquium Dinner/Evening Program supported by National Science Foundation**

Session Chair: Jonathan Furner

Dinner for colloquium members followed by additional activities.

- Covell Commons, West Coast Room & South Bay Room

Thursday, February 28th

8:00 am – noon **Conference Registration**
Participants will pick up their conference packets, final agenda, and name badges just outside of the Grand Horizon Room on the 3rd floor of Covell Commons.
• Covell Commons, Grand Horizon Room

9:00 – 11:30 am **Poster Sessions**
Session Chair: Jonathan Furner. Authors will present 30 posters related to new and promising research results and/or current design projects. Posters will remain available for viewing from 11:30 am to 5:30 pm. Posters must be taken down 5:30–5:45 pm on Thursday, February 28.
• Covell Commons, Northridge Room

noon – 1:30 pm **Conference Welcome Luncheon with Keynote Speaker Elizabeth D. Liddy**
UCLA Graduate School of Education & Information Studies Dean Aimée Dorr will welcome attendees to this year’s *iSchool* conference. Michael Eisenberg, from the University of Washington, will introduce the keynote speaker, Elizabeth D. Liddy, Dean of the School of Information Studies at Syracuse University.
• Covell Commons, Grand Horizon Room

1:30 – 2:00 pm **Afternoon Break :: Covell Commons, Grand Horizon Terrace**

2:00 – 3:30 pm **Early Afternoon Breakout Sessions**
Topics will range from information technologies and networking to information law and preservation.

➤ **PAPER SESSION #1: SESSION CHAIR: JOHN UNSWORTH • De Neve Plaza B**

Discipline Formation and the Field of Information	One iSchool’s Ideas and Identity: Doctoral Training and Research at Rutgers–SCILS 1959–2007	iSchools: Mice Roaring or the Future is Now Arriving?	iSchool Curricula: How Wide? How Deep?
---	---	---	--

➤ **PAPER SESSION #2: SESSION CHAIR: JENNY PREECE • Bradley Room 217**

Indigenous Peoples’ Rights to Culture and Individual Rights to Access	Tracing Transnational Flows of IT Knowledge Through Open Exchange of Software Development Know-how	Prim Drift, CopyBots, and Folk Preservation: Three Copyright Parables about Art in the Digital Age	Culture at the Interface: Collaborative Design and Information Management with Indigenous Communities
---	--	--	---

➤ **WILDCARD SESSIONS #1 – #4:**

(WC1) Is There a Cloud in Your Future? Applications of “Cloud Computing” to Web-scale Problems. • Covell Commons, West Coast Room	(WC2) Author Meets Critics: Sandra Braman. • Covell Commons, South Bay Room	(WC3) How Public Is the Internet? A Conversation on the Nature of Human Interactions Online and the Implications for Research Methods. • De Neve Plaza, Sycamore Room	(WC4) Narratives of Engagement: Working at the Intersections of Information, Social, and Domain Science. • Bradley Room 213/214
--	--	--	--

3:30 – 4:00 pm **Afternoon Break :: Covell Commons, Grand Horizon Terrace**

4:00 – 5:30 pm **Junior Faculty Mentoring :: Bradley Room 213/214**
Session Chairs: Leah A. Lievrouw and Steve Sawyer.
By invitation only.

4:00 – 5:30 pm Late Afternoon Breakout Sessions

Sessions continue to explore the conference themes and beyond.

► PAPER SESSION #3: SESSION CHAIR: BRIAN CANTWELL SMITH • De Neve Plaza B

The Role of the Arts in an iSchool Education	Community Identity: Peer Prestige and Academic Hiring in the iSchools	"Just Me in a Closet with a Computer": Student Perceptions of Undergraduate Information Technology Programs	Taylor's Value-added Model: Still Relevant After All These Years
--	---	---	--

► WILDCARD SESSIONS: #5 – #8:

(WC5) Bridging Discourses: Exploring the Relationship between Information Technologies and International Development • Covell Commons, South Bay Room	(WC6) iSchools and Undergraduate Education • Covell Commons, West Coast Room	(WC7) iSchools in Central and South Europe: Developments and Challenges of Cooperation • De Neve Plaza, Sycamore Room	(WC8) Can iSchools Fulfill Their Mission without Archival Studies? • Bradley Room 217
--	---	--	--

5:30 – 7:00 pm Microsoft® Reception

Join colleagues, mentors, and researchers on the Grand Horizon Terrace overlooking the UCLA campus. Refreshments and hors d'oeuvres will be served.

- Covell Commons, Grand Horizon Terrace

7:00 pm Dinner on your own

UCLA buses will transport you to your hotel or to one of more than a dozen restaurants in Westwood. Or make an optional visit to UCLA's Hammer Museum with an outstanding collection of both contemporary and traditional art. The return trip to your hotel from a restaurant or the Hammer will be on your own.

Friday, February 29th

8:00 – 8:30 am Continental Breakfast :: Covell Commons, Grand Horizon Terrace

8:30 – 10:00 am Early Morning Concurrent Sessions (continues on next page)

► PAPER SESSION #4: SESSION CHAIR: MARCIA BATES • Covell Commons, Northridge Room

Privacy and Participation in Ubiquitous Information Systems: Information Ethics when Mobile Phones are Sensors	Technical Authorship: Refiguring the Designer-user Conflict and the Visioning of Collective Technical Futures	The d-School in the i-School: HCI and Design Research	In the Spirit of Collaborating
--	---	---	--------------------------------

► PAPER SESSION #5: SESSION CHAIR: OLIVIA FROST • Covell Commons, South Bay Room

Understanding del.icio.us Tag Choice Using Simulations	Approaches to Semantic Markup for Natural Heritage Literatures	Reframing Book Publishing in the Age of Networking
--	--	--

► PAPER SESSION #6: SESSION CHAIR: STEPHEN HIRTLE • Covell Commons, West Coast Room

An Exploration of the Use of Games in Virtual Worlds for Online Education	Information Practices to Sustain Knowledge Building: The Case of the Virtual Math Teams Online Community	Designing e-Learning Games on Cellphones to Promote Language Learning and Literacy in the Developing World	Information Dissemination and Information Assurance in Vehicular Networks
---	--	--	---

Friday, February 29th

8:30 – 10:00 am Early Morning Concurrent Sessions (Continued)

► ROUNDTABLE SESSIONS #1 – #5:

(RT1) The Globalization of Education for Digital Librarianship—Implications for iSchools in North America • Covell Commons, Grand Horizon Room B	(RT2) Intersectionality and Interdisciplinarity: Information Studies and Studies of the “Other” • Covell Commons, Grand Horizon Room C	(RT3) Mediating Intimacy: Navigating Friendships, Family Ties, and Romantic Relationships with Communication Technologies • Covell Commons, Grand Horizon Room D
(RT4) Information Assurance, Intelligence and Security: Opportunities and Directions for Future Research • Covell Commons, Grand Horizon Room E	(RT5) Beyond the Qualitative/Quantitative Split: Alternate Forms of Research in the Information Space • Covell Commons, Grand Horizon Room F	

► WILDCARD SESSIONS #9 – #10:

(WC9) Research Perspectives on Social Tagging • De Neve Plaza B	(WC10) iSchool Proposal for Themed Wildcard Session on New Information Systems Methods • De Neve Plaza A
--	---

10:00 – 10:30 am Morning Break :: Covell Commons, Grand Horizon Terrace

10:30 – Noon Late Morning Concurrent Sessions

► PAPER SESSION #7: SESSION CHAIR: HARRY BRUCE • De Neve Plaza B

Reconsidering the Practical Application of Facets on the Web	The Social (Tagging) Act of Reading	FRBR as an Interdisciplinary High-middle Range Theory
--	-------------------------------------	---

► ROUNDTABLE SESSIONS #6 – #10:

(RT6) Whose Turn is It? Research on Gaming in the iSchools • Covell Commons, Grand Horizon Room B	(RT7) Schools in the Greater Community • Covell Commons, Grand Horizon Room C	(RT8) Young People and Technologies: Fostering Transformative Experiences • Covell Commons, Grand Horizon Room D
(RT9) Bridging the Divide Between Theory and Practice in ICT for Development • Covell Commons, Grand Horizon Room E	(RT10) Information: Transforming the World through Better Communications • Covell Commons, Grand Horizon Room F	

► WILDCARD SESSIONS #11 – #14:

(WC11) CHI, ICIS, SIGIR and Numerous Other Outlets That Won't Accept My Work: The Perils of Publishing Multidisciplinary Research? • Covell Commons, Northridge Room	(WC12) Mapping the Design Space of Design Education in iSchools • Covell Commons, West Coast Room	(WC13) Let's Get Wild: Building a National Research and Service Agenda for Community Technologies and Networking • Covell Commons, South Bay Room	(WC14) Info-scavengers • De Neve Plaza A
---	--	--	---

Noon – 1:30 pm Lunch :: Sunset Village Dining Hall

Noon – 1:30 pm Junior Faculty Mentoring and Lunch :: De Neve Plaza A *Session Chairs: Leah A. Lievrouw & Steve Sawyer.* By invitation only.

Noon – 1:30 pm Doctoral Colloquium Group Lunch :: Sunset Village Dining Hall By invitation only.

1:30 – 3:00 pm Early Afternoon Concurrent Sessions

➤ PAPER SESSION #8: SESSION CHAIR: DEBRA RICHARDSON • Covell Commons, Northridge Room

Ethics and the Undergraduate

➤ PAPER SESSION #9: SESSION CHAIR: BRUCE KINGMA • Covell Commons, West Coast Room

Community Informatics

➤ PAPER SESSION #10: SESSION CHAIR: DOUG OARD • De Neve Plaza B

Whose Data Do You Trust?
Integrity Issues in the
Preservation of Scientific Data

The Role of e-Infrastructures in
the Transformation of Research
Practices and Outcomes

Event-Driven
Productivity
Infrastructure

An Exploration of the
Life Cycle of eScience
Collaboratory Data

➤ ROUNDTABLE SESSIONS #11 – #14:

(RT11) Science and
Technology Studies
(STS) in iSchools
• Covell Commons,
Grand Horizon Room C

(RT12) The Role of
iSchools in Medical
Informatics
• Covell Commons,
Grand Horizon Room E

(RT13) Social
Network Analysis
• Covell Commons,
Grand Horizon
Room G

(RT14) Training, Integration,
and Identity: A Roundtable
Discussion of Undergraduate
and Professional Master's
Programs in iSchools
• Covell Commons,
Grand Horizon Room B

➤ WILDCARD SESSIONS #15 – #17:

(WC15) Okay, Facebook Me:
Exploring Behavior, Motivations
and Uses in Social Network Sites
• De Neve Plaza A

(WC16) What to Do with Geeks and
Nerds? A Collaborative Website
Analysis Workshop
• Covell Commons,
Grand Horizon Room F

(WC17) Information Credibility: Let's
Get Serious (and Have Some Fun Too!)
• Covell Commons, South Bay Room

3:00 pm The Getty Center

Board buses for a visit to the Getty Research Institute and an evening at the Getty Museum. This special event features a program from Getty Research Institute (GRI) speakers, followed by a reception. Conference participants may dine at the Café or Restaurant (if you have reservations for the Restaurant) on the Getty campus, visit the GRI and/or Museum galleries until 9 pm.

Saturday, March 1st

8:00 – 9:00 am Continental Breakfast :: Northwest Campus Auditorium Patio

9:00 – 10:30 am Keynote Presentation by Christa Phillips, Xbox.com, Microsoft® Corporation
• Northwest Campus Auditorium

10:30 – 11:00 am Morning Break :: Northwest Campus Auditorium Patio

11:00 – Noon Plenary Concluding Session :: Northwest Campus Auditorium

Poster Sessions

SESSION CHAIR: Jonathan Furner • TIME: Thursday 9:00 am • LOCATION: Northridge Room

POSTER #1

Federal Funding for Information Research

Trudi Bellardo Hahn

University of Maryland; thahn@umd.edu

Topics: nature and scope of iSchools and iResearch

Keywords: federal funding, research

POSTER #2

From Cultural Participation to Information Visualization: A New Framework for Old Knowledge Management Schemas

Theodore Patrick Milas

Florida State University; pmilas@fsu.edu

Topics: cultural information systems, information organization, information technology and services for under-addressed communities, nature and scope of iSchools and iResearch

Keywords: knowledge management, non-Western social informatics

POSTER #3

"Community Innovativeness": A New Perspective on Knowledge Creation

Benjamin Addom

Syracuse University; bkaddom@syr.edu

Topics: community technologies and networking, information technology and services for under-addressed communities

Keywords: community informatics, innovation, knowledge generation

POSTER #4

Why Leave Wikipedia?

Lian Jian and Jeffrey K. MacKie-Mason

University of Michigan, Ann Arbor; ljian@umich.edu

Topics: community technologies and networking, information organization, information management, information policy, ethics and law

Keywords: user-contributed-content community management systems

POSTER #5

Beyond the Digital Divide into Computer-Mediated Communications: A Content Analysis of the Role of Community Weblogs in Building Oldenburg's Virtual Third Places in Black America

Frank Igwe

Pennsylvania State University; frankigwe@hotmail.com

Topics: community technologies and networking, cultural information systems, information technology and services for under-addressed communities

Keywords: HIV/AIDS, blogs, community, third place, social support, content analysis, African-American

POSTER #6

Older Adults and Information Technology: The Current State of Research and Future Directions

Johanna L. H. Birkland and Michelle L. Kaarst-Brown

Syracuse University; jlbirkla@syr.edu

Topics: information technology and services for under-addressed communities, other

Keywords: older adults, seniors, elderly, information technology

POSTER #7

A Preliminary Consumer Health Information-seeking (CHI-seeking) Behavior Model of Physicians Who Treat Elderly Depressed Patients (Results from the Pilot Study of a Dissertation Proposal)

Mary Jo Dorsey

University of Pittsburgh; mjd21@pitt.edu

Topics: information technology and services for under-addressed communities

Keywords: information-seeking behavior, consumer health information, models, physicians, seniors

POSTER #8

Developing an Organizational Model and Technical Implementation Plan for Wireless Mesh Networks in Sao Tome and Principe, West Africa

Christopher J. Ritzo

University of Illinois at Urbana-Champaign; critz1@uiuc.edu

Topics: community technologies and networking, information infrastructure development, information technology and services for under-addressed communities

Keywords: wireless mesh networks, community informatics, digital divide, information society, Sao Tome and Principe

POSTER #9

Interactive Machine Learning (IML) Markup of OCR Generated Text by Exploiting Domain Knowledge: A Biodiversity Case Study

Qin Wei and P. Bryan Heidorn

University of Illinois at Urbana-Champaign; qinwei2@uiuc.edu

Topics: community technologies and networking, information infrastructure development, other

Keywords: Interactive Machine Learning, biodiversity informatics, evaluation of IML, information extraction, automatic markup

POSTER #10

Post-disaster Information Infrastructure: The 1989 Loma Prieta Earthquake

Megan Finn

University of California, Berkeley; megfinn@ischool.berkeley.edu

Topics: community technologies and networking, information infrastructure development, other

Keywords: disasters; information infrastructure; information ecologies; social studies of information; sociology of information

POSTER #11

Towards a Data and Workflow Collaboratory for Research on Free and Open Source Software and its Development

James Howison, Kevin Crowston, and Megan Conklin

Syracuse University; Elon University; jhowison@syr.edu

Topics: information infrastructure development, preserving digital information

Keywords: cyberinfrastructure, repositories

POSTER #12

Towards a Model of Determinants of Web Services Platform Adoption by Complementers

Joseph B. Rubleske

Syracuse University; jrubleske@gmail.com

Topics: information infrastructure development

Keywords: software platforms, web services, adoption, network effects, complementarities

POSTER #13

ChemXSeer: An eScience Web Search Engine and Repository for Chemistry

C. Lee Giles, Prasenjit Mitra, Karl Muller, James Kubicki, Barbara Garrison, James Z. Wang, Bingjun Sun, Ying Liu, Qingzhao Tan, Levent Bolelli, Xiaonan Lu, Anuj Jaiswal, Kun Bai, Isaac Council, William Brouwer, Juan Fernandez, and Joel Bandstra
Pennsylvania State University; giles@ist.psu.edu

Topics: community technologies and networking, information infrastructure development, information organization, preserving digital information

Keywords: e-science, chemistry, cyberinfrastructure, information retrieval, data extraction

POSTER #14

Stability and Change in Self-organizing Technology-supported Groups: Evidence from Genres of Communication in Free and Open Source Software Development

James Howison

Syracuse University; jhowison@syr.edu

Topics: community technologies and networking, other

Keywords: genres of online communications, distributed teams, organizational change

POSTER #15

The Impacts of Gender and Initial Social Activities on Trust and Performance in CMC

Xiaoning Sun

Drexel University;

xiaoning.sun@ischool.drexel.edu

Topics: other

Keywords: gender, computer-mediated communication, trust, Instant Messaging

POSTER #16

Healthcare Informatics: Supporting Collaborative Sensemaking in the Emergency Department

Sharoda A. Paul

Pennsylvania State University; spaul@ist.psu.edu

Topics: other

Keywords: healthcare informatics, ICTs in healthcare, computer-supported cooperative work, collaborative systems, sense-making

POSTER #17

Support Sense-making with Tools for Structuring a Conceptual Space

Pengyi Zhang

University of Maryland; pengyi@umd.edu

Topics: information organization, information management

Keywords: sense-making tools, conceptual structure, user-centered design and evaluation

POSTER #18

Deriving Ontology from Folksonomy and Controlled Vocabulary

Miao Chen and Jian Qin

Syracuse University; mchen14@syr.edu

Topics: information organization

Keywords: ontology, folksonomy, tag, controlled vocabulary, vector space

POSTER #19

The Impact of Documentation on Secondary Data Use

Jinfang Niu

University of Michigan; niujf@umich.edu

Topics: information organization

Keywords: secondary data use, documentation, data sharing, metadata

POSTER #20

Proposal of Document Classification with Word Sense Disambiguation

Xiaozhong Liu

Syracuse University; xliu12@syr.edu

Topics: information organization, other

Keywords: Document classification, WSD, NLP

POSTER #21

Group Maintenance Behaviors in Self-organizing Distributed Teams

Na Li, Michael John Scialdone, Robert Heckman, and Kevin Crowston

Syracuse University; nli@syr.edu

Topics: community technologies and networking

Keywords: group maintenance, self-organizing distributed team, FLOSS, politeness theory, content analysis

POSTER #22

Longitudinal Analysis of Freshman Adoption of Facebook.com

Frederic Stutzman

University of North Carolina-Chapel Hill;

fstutzman@gmail.com

Topics: community technologies and networking, other

Keywords: Facebook, social networks, identity, friendship, data

POSTER #23

Cyber Java Monopoly: Game-based Approach of Collaborative Programming Language Learning

I-Han Hsiao and Yi-Ling Lin

University of Pittsburgh; ihh4@pitt.edu

Topics: community technologies and networking

Keywords: game-based learning, collaborative learning, programming language learning

POSTER #24

The Collaborative Information Behavior of Middle School Students in Online Learning Environments: An Exploratory Study

Nan Zhou and Denise Agosto

Drexel University;

nan.zhou@ischool.drexel.edu

Topics: community technologies and networking, Other

Keywords: collaborative information behavior, computer-supported collaborative learning

POSTER #25

A Visual Map of Public Mementos and Conjectures

Alberto Pepe

University of California, Los Angeles; apepe@ucla.edu

Topics: cultural information systems, information organization, information policy, ethics and law

Keywords: information visualization, mood analysis, data mining

POSTER #26

To Catch a Predator? The MySpace Moral Panic

Alice E. Marwick

New York University;

alice.marwick@nyu.edu

Topics: community technologies and networking, information policy, ethics and law

Keywords: social networking, information policy, moral panics, internet filtering

POSTER #27

Conversation Repository for Participatory Librarianship

Keisuke Inoue

Syracuse University; kinoue@syr.edu

Topics: Community technologies and networking, information organization

Keywords: information seeking behavior, information retrieval, Web 2.0, discourse analysis, conversation theory

POSTER #28

Watching Organizational Opinion via Social Tagging

Terrell Russell

University of North Carolina-Chapel Hill; unc@terrellrussell.com

Topics: community technologies and networking, information organization, information management

Keywords: expertise, organizations, tagging, folksonomy

POSTER #29

Behavioral Parameters of Trustworthiness for Countering Insider Threats

Shuyuan Mary Ho

Syracuse University; smho@syr.edu

Topics: information assurance and security

Keywords: trustworthiness, insider threats, personnel anomaly detection

POSTER 30#

Incentive Design for Home Computer Security

Rick Wash and Jeffrey King MacKie-Mason

University of Michigan; rwash@umich.edu

Topics: information assurance and security

Keywords: security incentives economics design

Paper Sessions

PAPER SESSION #1

Session Chair: John Unsworth
Thursday, 2:00 – 3:30 pm
Location: De Neve Plaza B

Discipline Formation and the Field of Information

Fran Miksa

University of Texas at Austin;
miksa@ischool.utexas.edu

Topics: information organization, nature and scope of iSchools and iResearch

Keywords: discipline formation, tradition, information organization, field of information

One iSchool's Ideas and Identity: Doctoral Training and Research at Rutgers–SCILS 1959–2007

Hartmut B. Mokros

Rutgers University;
mokros@scils.rutgers.edu

Topics: nature and scope of iSchools and iResearch

Keywords: education, doctoral training, research, iSchools

iSchools: Mice Roaring or the Future is Now Arriving?

Steve Sawyer and Howard Rosenbaum
Pennsylvania State University;
Indiana University; sawyer@ist.psu.edu

Topics: nature and scope of iSchools and iResearch, other

Keywords: iSchools, institutions, faculty, graduate students

iSchool Curricula: How Wide? How Deep?

Richard A. Thompson
University of Pittsburgh;
thompson@sis.pitt.edu

Topics: nature and scope of iSchools and iResearch

Keywords: curriculum, telecommunications

PAPER SESSION #2

Session Chair: Jenny Preece
Thursday, 2:00 – 3:30 pm
Location: Bradley Rm 217

Indigenous Peoples' Rights to Culture and Individual Rights to Access

Kristy Kay Mathiesen

University of Arizona;
kmathies@email.arizona.edu

Topics: information policy, ethics and law

Keywords: human rights, cultural property, indigenous, philosophy, intellectual freedom

Tracing Transnational Flows of IT Knowledge Through Open Exchange of Software Development Know-How

Susan Elliott Sim and Kavita Philip

University of California, Irvine;
ses@ics.uci.edu

Topics: information policy, ethics and law, information technology and services for under-addressed communities

Keywords: software engineering, India, ethnography, history

Prim Drift, CopyBots, and Folk Preservation: Three Copyright Parables about Art in the Digital Age

Kari M. Kraus

University of Maryland;
karimkraus@gmail.com

Topics: cultural information systems, information policy, ethics and law, other

Keywords: second life, virtual worlds, intellectual property, arts and humanities

Culture At the Interface: Collaborative Design and Information Management with Indigenous Communities

Kimberly Christen, Craig Dietrich, and Ramesh Srinivasan

Washington State University; University of Iowa; University of California, Los Angeles; kim.christen@gmail.com

Topics: cultural information systems, information technology and services for under-addressed communities

Keywords: indigenous, digital archives, information management, collaboration

PAPER SESSION #3

Session Chair: Brian Cantwell Smith
Thursday, 4:00 – 5:30 pm
Location: De Neve Plaza B

The Role of the Arts in an iSchool Education

Robert Heckman and Jaime Snyder

Syracuse University; rheckman@syr.edu

Topics: nature and scope of iSchools and iResearch, other

Keywords: arts, aesthetics, iprofessionals

Community Identity: Peer Prestige and Academic Hiring in the iSchools

Andrea Wiggins, Michael J. McQuaid, and Lada A. Adamic

Syracuse University;
University of Michigan; awiggins@syr.edu

Topics: nature and scope of iSchools and iResearch

Keywords: identity, community, prestige, hiring

"Just Me in a Closet with a Computer": Student Perceptions of Undergraduate Information Technology Programs

Heekyung Choi

University of Illinois at Urbana-Champaign; hchoi8.uiuc@gmail.com

Topics: nature and scope of iSchools and iResearch

Keywords: informatics, IT education

Taylor's Value-Added Model: Still Relevant After All These Years

Michael B. Eisenberg and Lee Dirks

University of Washington;
Microsoft Corp; mbe@u.washington.edu

Topics: nature and scope of iSchools and iResearch

Keywords: value-added, information theory, design, information technology, systems

PAPER SESSION #4

Session Chair: Marcia Bates
Friday, 8:30 – 10:00 am
Location: Covell Commons, Northridge Rm

Privacy and Participation in Ubiquitous Information Systems: Information Ethics when Mobile Phones are Sensors

Katie Shilton, Jeff Burke, Deborah Estrin, Mark Hansen and Mani Srivastava

University of California, Los Angeles;
kshilton@ucla.edu

Topics: information assurance and security, information policy, ethics and law, information technology and services for under-addressed communities

Keywords: ubiquitous computing, privacy, ethics, embedded networked sensing

Technical Authorship: Refiguring the Designer-user Conflict and the Visioning of Collective Technical Futures

Marisa Leavitt Cohn, Susan Elliott Sim and Kavita Philip

University of California, Irvine;
mlcohn@ics.uci.edu

Topics: information technology and services for under-addressed communities

Keywords: authorship, user-centered design, STS, software

The d-School in the i-School: HCI and Design Research

John M. Carroll, S. R. Haynes, F. E. Ritter, M. B. Rosson and X. L. Zhang

Pennsylvania State University;
jmcarroll@psu.edu

Topics: other

Keywords: i-School, d-School, identity, interdisciplinarity, human-computer interaction

In the Spirit of Collaborating

Michael Nilan and Michael D'Eredita

Syracuse University; mnilan@syr.edu

Topics: other

Keywords: cognitive, social cognition, collaborating, communicating, information use

PAPER SESSION #5

Session Chair: Olivia Frost
Friday, 8:30 – 10:00 am
Location: Covell Commons, South Bay Rm

Understanding del.icio.us Tag Choice Using Simulations

Rick Wash and Emilee Rader

University of Michigan; rwash@umich.edu

Topics: information organization

Keywords: delicious tagging KM

Approaches to Semantic Markup for Natural Heritage Literatures

Hong Cui
University of Arizona;
hongcui@email.arizona.edu

Topics: community technologies and networking, information organization, information management, information technology and services for under-addressed communities

Keywords: natural heritage literature, semantic markup, taxonomists, machine learning

Reframing Book Publishing in the Age of Networking

Jana Bradley and Heather Vokac
University of Arizona;
janabrad@email.arizona.edu

Topics: community technologies and networking, cultural information systems, information technology and services for under-addressed communities

Keywords: electronic publishing, emerging publishers, publishing in the 21st century

PAPER SESSION #6

Session Chair: Stephen Hirtle
Friday, 8:30 – 10:00 am
Location: Covell Commons, West Coast Rm

An Exploration of the Use of Games in Virtual Worlds for Online Education

Martha Garcia-Murillo and Ian MacInnes
Syracuse University; mgarciam@syr.edu

Topics: other

Keywords: online education, virtual worlds, Gee's learning principles

Information Practices to Sustain Knowledge Building: The Case of the Virtual Math Teams Online Community

Johann Sarmiento and Gerry Stahl
Drexel University; jsarmi@drexel.edu

Topics: community technologies and networking, information infrastructure development, other

Keywords: information practices, social computing, knowledge building, computer-supported collaborative learning

Designing e-Learning Games on Cellphones to Promote Language Learning and Literacy in the Developing World

Matthew Kam
University of California, Berkeley;
mattkam@cs.berkeley.edu

Topics: cultural information systems, information technology and services for under-addressed communities

Keywords: cellphone, developing world, games, e-learning, literacy,

Information Dissemination and Information Assurance in Vehicular Networks

Prashant Krishnamurthy
University of Pittsburgh; prashk@pitt.edu

Topics: community technologies and networking, information assurance and security

Keywords: vehicular network, community network, information assurance

PAPER SESSION #7

Session Chair: Harry Bruce
Friday, 10:30 am – noon
Location: De Neve Plaza B

Reconsidering the Practical Application of Facets on the Web

Kathryn La Barre
University of Illinois at Urbana-Champaign; kathryn.labarre@gmail.com

Topics: information organization

Keywords: faceted analysis, metadata, resource description

The Social (Tagging) Act of Reading

Cheryl Knott Malone
University of Arizona;
ckmalone@u.arizona.edu

Topics: information organization

Keywords: social tagging, metadata

FRBR as an Interdisciplinary High-middle Range Theory

Allen H. Renear, David S. Dubin, and Cheryl Malone

University of Illinois at Urbana-Champaign; University of Arizona;
renear@uiuc.edu

Topics: information organization

Keywords: metadata, resource description, FRBR

PAPER SESSION #8

Session Chair: Debra Richardson
Friday, 1:30 – 3:00 pm
Location: Covell Commons, Northridge Rm

Ethics and the Undergraduate

Paul Conway, Toni Carbo, and Lynette Kvasny
University of Michigan;
University of Pittsburgh;
Pennsylvania State University;
pconway@umich.edu

Topics: information policy, ethics and law

Keywords: information ethics, undergraduate education, pedagogy, information policy, social justice

PAPER SESSION #9

Session Chair: Bruce Kingma,
Friday, 1:30 – 3:00 pm
Location: Covell Commons, West Coast Rm

Community Informatics

Kate Williams, Joan Durrance, Howard Rosenbaum, and John Carroll
University of Illinois;
University of Michigan;
Indiana University; katewill@uiuc.edu

Topics: community technologies and networking, information infrastructure development, information technology and services for under-addressed communities, nature and scope of iSchools and iResearch

Keywords: community, community informatics, memory, datasets, outcomes, curriculum

PAPER SESSION #10

Session Chair: Doug Oard
Friday, 1:30 – 3:00 pm
Location: De Neve Plaza B

Whose Data Do You Trust? Integrity Issues in the Preservation of Scientific Data

Matthew S. Mayernik, Jillian C. Wallis, Alberto Pepe, and Christine L. Borgman
University of California, Los Angeles;
mattmayernik@ucla.edu

Topics: information assurance and security, information infrastructure development, preserving digital information

Keywords: data science, data integrity, data curation, cyberinfrastructure, data reuse

The Role of e-Infrastructures in the Transformation of Research Practices and Outcomes

Eric T. Meyer, Ralph Schroeder, and William H. Dutton
University of Oxford;
eric.meyer@oii.ox.ac.uk

Topics: information infrastructure development, information policy, ethics and law

Keywords: e-Research, cyberinfrastructure, e-Social Science, social shaping

Event-Driven Productivity Infrastructure

Hugh Taylor
Microsoft Corp; htaylor@microsoft.com

Topics: information organization, information management

Keywords: EDA, Event-Driven Architecture, Productivity Infrastructure

An Exploration of the Life Cycle of eScience Collaboratory Data

Jillian C. Wallis, Matthew S. Mayernik, Alberto Pepe, and Christine L. Borgman
University of California, Los Angeles;
jwallisi@ucla.edu

Topics: information infrastructure development, information management

Keywords: data life cycle, cyberinfrastructure

Wildcard & Roundtable Sessions

Wildcard

WILDCARD SESSION #1

Session Chair: Jimmy Lin
Thursday, 2:00 – 3:30 pm
Location: Covell Commons, West Coast Rm

Is There a Cloud in your Future? Applications of “Cloud Computing” to Web-scale Problems

Jimmy Lin
University of Maryland; jimmylin@umd.edu

Topics: information infrastructure development, information technology and services for under-addressed communities, nature and scope of iSchools and iResearch

Keywords: cloud computing

WILDCARD SESSION #2

Session Chair: Jean-François Blanchette
Thursday, 2:00 – 3:30 pm
Location: Covell Commons, South Bay Rm

Author Meets Critics: Sandra Braman

Jean-François Blanchette, Sandra Braman, Steve Jackson, Leah Lievrouw, and Milton Mueller

University of California, Los Angeles;
University of Wisconsin-Milwaukee;
University of Michigan;
Syracuse University; blanchette@ucla.edu

Topics: information policy, ethics and law

Keywords: information policy, law, author meets critics, inter-disciplinarity, Sandra Braman

WILDCARD SESSION #3

Session Chair: Alpha Selene DeLap
Thursday, 2:00 – 3:30 pm
Location: Covell Commons, Sycamore Rm

How Public Is the Internet? A Conversation on the Nature of Human Interactions On-line and the Implications for Research Methods

Alpha Selene DeLap and Phillip M. Edwards
University of Washington;
adelap@u.washington.edu

Topics: information policy, ethics and law,

Keywords: ethics, public and private uses of the Internet, Web 2.0 technologies, IRB, online research

WILDCARD SESSION #4

Session Chair: David Ribes
Thursday, 2:00 – 3:30 pm
Location: Bradley Rm 213/214

Narratives of Engagement: Working at the Intersections of Information, Social, and Domain Science

David Ribes, Christine Borgman, Thomas Finholt, Karen Baker, Geoffrey Bowker
University of Michigan; dribes@umich.edu

Topics: nature and scope of iSchools and iResearch, other

Keywords: collaboration, engagement, social/information science

WILDCARD SESSION #5

Session Chair: Ramesh Srinivasan
Thursday, 4:00 – 5:30 pm
Location: Covell Commons, South Bay Rm

Bridging Discourses: Exploring the Relationship between Information Technologies and International Development

Ramesh Srinivasan, Steven Jackson, Tapan Parikh, and Jessica Wallack
University of California, Los Angeles;
University of Michigan;
University of California, Berkeley;
University of California, San Diego;
Center for Development Finance, India;
srinivasan@ucla.edu

Topics: community technologies and networking, cultural information systems, information technology and services for under-addressed communities

Keywords: development, international, information society, microfinance, community

WILDCARD SESSION #6

Session Chair: Derek L. Hansen
Thursday, 4:00 – 5:30 pm
Location: Covell Commons, West Coast Rm

iSchools and Undergraduate Education

Derek L. Hansen, Susan M. Bonzi, Lori Kendall, and David W. McDonald
University of Maryland;
Syracuse University; University of Illinois at Urbana-Champaign;
University of Washington;
shakmatt@gmail.com

Topics: nature and scope of iSchools and iResearch

Keywords: undergraduate education

WILDCARD SESSION #7

Session Chair: Tatjana Aparac-Jelusic
Thursday, 4:00 – 5:30 pm
Location: De Neve Plaza, Sycamore Rm

iSchools in Central and South Europe: Developments and Challenges of Cooperation

Tatjana Aparac-Jelusic, Christian Schloegl, Maja Zumer and Anna Maria Tamamro
University of Zadar, Croatia (Hrvatska);
taparac@unizd.hr

Topics: nature and scope of iSchools and iResearch

Keywords: ischools, Central and South Europe, education policy, curricula development, distance education, cooperation in education

WILDCARD SESSION #8

Session Chair: Richard James Cox
Thursday, 4:00 – 5:30 pm
Location: Bradley Rm 217

Can iSchools Fulfill Their Mission without Archival Studies?

Richard James Cox, Ronald Larsen, Michèle V. Cloonan, and Anne J. Gilliland
University of Pittsburgh School;
Simmons College; University of California, Los Angeles; rjcox111@comcast.net

Topics: cultural information systems, preserving digital information, other

Keywords: archives, preservation, heritage, memory, curation

WILDCARD SESSION #9

Session Chair: Alla Zollers
Friday, 8:30 – 10:00 am
Location: De Neve Plaza B

Research Perspectives on Social Tagging

Alla Zollers, Tony Moore, Lilly Nguyen, Terrell Russell, and Fred Stutzman
University of California, Los Angeles;
University of North Carolina-Chapel Hill;
alla.zollers@ucla.edu

Topics: Community technologies and networking, information organization

Keywords: tags, social tagging, social classification, social software

WILDCARD SESSION #10

Session Chair: Susan Leigh Star
Friday, 8:30 – 10:00 am
Location: De Neve Plaza A

iSchool Proposal for Themed Wildcard Session on New Information Systems Methods

Susan Leigh Star, Marina Jirotko, Jack Whalen, and Scout Calvert
Santa Clara University; Oxford University;
Xerox PARC; University of California, Santa Cruz; slstar@scu.edu

Topics: nature and scope of iSchools and iResearch

Keywords: methods, assessment, STS

WILDCARD SESSION #11

Session Chair: Jim Jansen
Friday, 10:30 am – noon
Location: Covell Commons, Northridge Rm

CHI, ICIS, SIGIR and Numerous Other Outlets That Won't Accept My Work: The Perils of Publishing Multidisciplinary Research?

Jim Jansen, Howard Rosenbaum, Mark Ackerman, and Elizabeth D. Liddy
Pennsylvania State University;
Indiana University; University of Michigan;
Syracuse University; jjansen@acm.org

Topics: nature and scope of iSchools and iResearch, other

Keywords: publishing, multidisciplinary research

WILDCARD SESSION #12

Session Chair: M. Cameron Jones
Friday, 10:30 am – noon
Location: Covell Commons, West Coast Rm

Mapping the Design Space of Design Education in iSchools

M. Cameron Jones, Ingbert R. Floyd, Michael B. Twidale, Piotr D. Adamczyk, and John Carroll
University of Illinois at Urbana-Champaign; Pennsylvania State University;
mjones2@uiuc.edu

Topics: nature and scope of iSchools and iResearch

Keywords: design, design thinking, education, pedagogy

WILDCARD SESSION #13

Session Chair: Michael D. Crandall
Friday, 10:30 am – noon
Location: Covell Commons, South Bay Rm

Let's Get Wild: Building a National Research and Service Agenda for Community Technologies and Networking

Michael D. Crandall, Karen Fisher, Karine Barzilai-Nahon, and Robert Mason
University of Washington;
mikecran@u.washington.edu

Topics: community technologies and networking, information technology and services for under-addressed communities

Keywords: community technologies, research, impact, society

WILDCARD SESSION #14

Session Chair: Daniela Karin Rosner
Friday, 10:30 am – noon
Location: De Neve Plaza A

Info-scavengers

Daniela Karin Rosner, Kevin Mateo Lim, and Hannes Hesse
University of California, Berkeley;
daniela@ischool.berkeley.edu

Topics: information organization, other

Keywords: immersive games, information retrieval, performance, alternative reality games

WILDCARD SESSION #15

Session Chair: Frederic D. Stutzman
Friday, 1:30 – 3:00 pm
Location: De Neve Plaza A

Okay, Facebook Me: Exploring Behavior, Motivations and Uses in Social Network Sites

Frederic D. Stutzman, Danah Boyd, Alice Marwick, and Cliff Lampe
University of North Carolina-Chapel Hill;
University of California, Berkeley;
New York University;
Michigan State University;
fred@metalab.unc.edu

Topics: community technologies and networking, other

Keywords: social networks, facebook, myspace, social computing, identity

WILDCARD SESSION #16

Session Chair: Indira Rita Guzman
Friday, 1:30 – 3:00 pm
Location: Covell Commons, Grand Horizon Rm F

What to Do with Geeks and Nerds? A Collaborative Website Analysis Workshop

Kathryn Stam, Indira Rita Guzman, and Michael Scialdone
SUNY Institute of Technology;
TUI University; Syracuse University;
stmkath@gmail.com

Topics: cultural information systems, other

Keywords: web research

WILDCARD SESSION #17

Session Chair: Michael B. Eisenberg
Friday, 1:30 – 3:00 pm
Location: Covell Commons, South Bay Rm

Information Credibility: Let's Get Serious (and Have Some Fun Too!)

Michael B. Eisenberg, Peyina Lin, Yuan Lin, R. David Lankes, Miriam Metzger, and Soo Young Rieh
University of Washington;
Syracuse University;
University of California, Santa Barbara;
University of Michigan;
mbe@u.washington.edu

Topics: community technologies and networking, information infrastructure development, information technology and services for under-addressed communities

Keywords: information credibility, social networks, systems development, credibility

Roundtable

ROUNDTABLE SESSION #1

Session Chair: Terry Weech
Friday, 8:30 – 10:00 am
Location: Covell Commons, Grand Horizon Rm B

The Globalization of Education for Digital Librarianship—Implications for iSchools in North America

Terry Weech, Heather Moulaison, Tatjana Aparac-Jelusic, Stephen Downie, Michael Lesk, and Anne Maria Tammaro
University of Illinois; Rutgers University;
University of Zadar, Croatia (Hrvatska);
University of Illinois at Urbana-Champaign; University of Parma, Italy;
weech@uiuc.edu

Topics: information policy, ethics and law, nature and scope of iSchools and iResearch

Keywords: digital libraries, iSchools, education, international

ROUNDTABLE SESSION #2

Session Chair: Clara M. Chu
Friday, 8:30 – 10:00 am
Location: Covell Commons, Grand Horizon Rm C

Intersectionality and Interdisciplinarity: Information Studies and Studies of the "Other"

Clara M. Chu
University of California, Los Angeles;
cchu@ucla.edu

Topics: nature and scope of iSchools and iResearch

Keywords: intersectionality, interdisciplinarity, research, education, the other

ROUNDTABLE SESSION #3

Session Chair: Andrew T. Fiore
Friday, 8:30 – 10:00 am
Location: Covell Commons, Grand Horizon Rm D

Mediating Intimacy: Navigating Friendships, Family Ties, and Romantic Relationships with Communication Technologies

Andrew T. Fiore, Coye Cheshire, and Christo Sims
University of California, Berkeley;
atf@ischool.berkeley.edu

Topics: community technologies and networking, cultural information systems

Keywords: intimacy, friendship, family, mediation, communication

Roundtable Sessions

ROUNDTABLE SESSION #4

Session Chair: Scott Paquette
Friday, 8:30 – 10:00 am
Location: Covell, Grand Horizon Rm E

Information Assurance, Intelligence and Security: Opportunities and Directions for Future Research

Scott Paquette, Isabelle Fagnot, Kevin C. Desouza, Dave Yates, and Shuyuan Mary Ho
University of Maryland;
Syracuse University;
University of Washington;
University of Southern California;
spaquette@umd.edu

Topics: information assurance and security, information management, mature and scope of iSchools and iResearch

Keywords: information security, assurance, intelligence

ROUNDTABLE SESSION #5

Session Chair: Deborah Turner
Friday, 8:30 – 10:00 am
Location: Covell, Grand Horizon Rm F

Beyond the Qualitative/Quantitative Split: Alternate Forms of Research in the Information Space

Deborah Turner, Melanie Feinberg, Lisa Nathan, and Kris Unsworth
University of Washington;
turned@u.washington.edu

Topics: nature and scope of iSchools and iResearch

Keywords: research, methods, information science

ROUNDTABLE SESSION #6

Session Chair: Scott Nicholson
Friday, 10:30 am – noon
Location: Covell, Grand Horizon Rm B

Whose Turn is It? Research on Gaming in the iSchools

Scott Nicholson and Ian MacInnes
Syracuse University; srnichol@syr.edu

Topics: community technologies and networking, information technology and services for under-addressed communities

Keywords: gaming

ROUNDTABLE SESSION #7

Session Chair: Eric Christopher Kansa
Friday, 10:30 am – noon
Location: Covell, Grand Horizon Rm C

iSchools in the Greater Community

Eric Christopher Kansa, Erik Wilde, and Raymond Yee
University of California, Berkeley;
ekansa@ischool.berkeley.edu

Topics: nature and scope of iSchools and iResearch

Keywords: iSchools, clinic, instruction, projects, collaboration, laboratory, incubation, university

ROUNDTABLE SESSION #8

Session Chair: Carol L. Tilley
Friday, 10:30 am – noon
Location: Covell, Grand Horizon Rm D

Young People and Technologies: Fostering Transformative Experiences

Carol L. Tilley
University of Illinois at Urbana-Champaign; ctiley@uiuc.edu

Topics: information technology and services for under-addressed communities

Keywords: young people, information technology, 21st century literacies, professional education

ROUNDTABLE SESSION #9

Session Chair: Melissa Rita Ho
Friday, 10:30 am – noon
Location: Covell, Grand Horizon Rm E

Bridging the Divide Between Theory and Practice in ICT for Development

Melissa Rita Ho and Rajesh Veeraraghavan
University of California, Berkeley;
mho@ischool.berkeley.edu

Topics: information technology and services for under-addressed communities, nature and scope of iSchools and iResearch

Keywords: ictd ict4d development emerging regions

ROUNDTABLE SESSION #10

Session Chair: Charles DuBois
Friday, 10:30 am – noon
Location: Covell, Grand Horizon Rm F

Information: Transforming the World through Better Communications

Charles DuBois, Ronald Dietel, Marlo Welshons, and Kelly Shaffer
Pennsylvania State University,
University of California, Los Angeles;
University of Illinois at Urbana-Champaign; University of Pittsburgh;
dietel@cse.ucla.edu

Topics: cultural information systems, information infrastructure development, information organization, nature and scope of iSchools and iResearch, other

Keywords: communications, mentoring, professional development, writing, speaking

ROUNDTABLE SESSION #11

Session Chair: Nancy Van House
Friday, 1:30 – 3:00 pm
Location: Covell, Grand Horizon Rm B

Science and Technology Studies (STS) in iSchools

Nancy Van House, Jean-François Blanchette, Christine L. Borgman, Leah Lievrouw, Susan Leigh Star, and Judy Weedman
University of California, Berkeley;
University of California, Los Angeles;
Santa Clara University;
San Jose State University;
vanhouse@ischool.berkeley.edu

Topics: nature and scope of iSchools and iResearch

Keywords: STS, methods, theory, IS

ROUNDTABLE SESSION #12

Session Chair: Saira Naim Haque
Friday, 1:30 – 3:00 pm
Location: Covell, Grand Horizon Rm E

The Role of iSchools in Medical Informatics

Saira Naim Haque, Carsten Svarr Oesterlund, and Madhu C. Reddy
Syracuse University; Pennsylvania State University; snhaque@syr.edu

Topics: information infrastructure development, information management, nature and scope of iSchools and iResearch

Keywords: health informatics; iSchool

ROUNDTABLE SESSION #13

Session Chair: John Yen
Friday, 1:30 – 3:00 pm
Location: Covell, Grand Horizon Rm G

Social Network Analysis

John Yen, Lee Giles, Hank Foley, Lada Adamic, Filippo Menczer, Haizheng Wang, Andrea Tapia, Marc Smith, Alessandro Vespigani, and Brian Smith
Pennsylvania State University;
University of Michigan; Indiana University;
Microsoft Corp; jyen@ist.psu.edu

Topics: community technologies and networking

Keywords: social network; community; network analysis

ROUNDTABLE SESSION #14

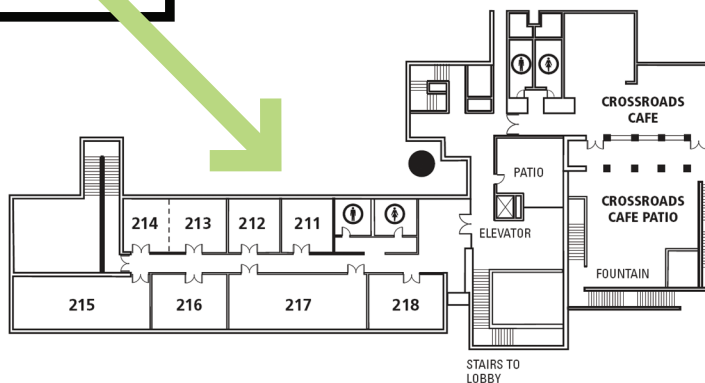
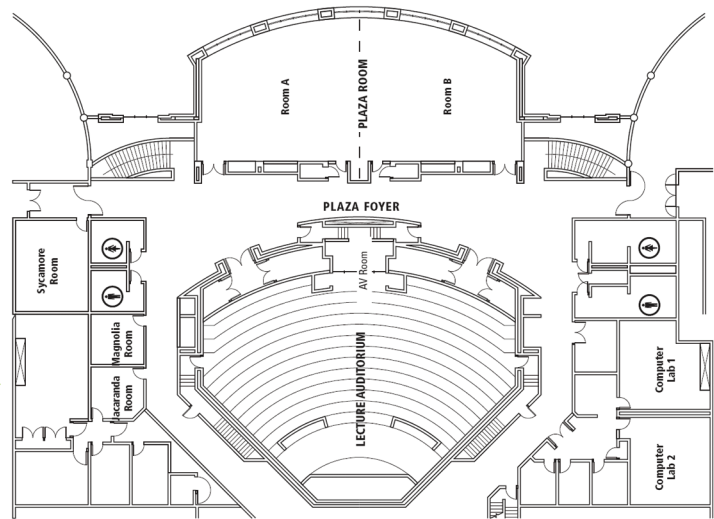
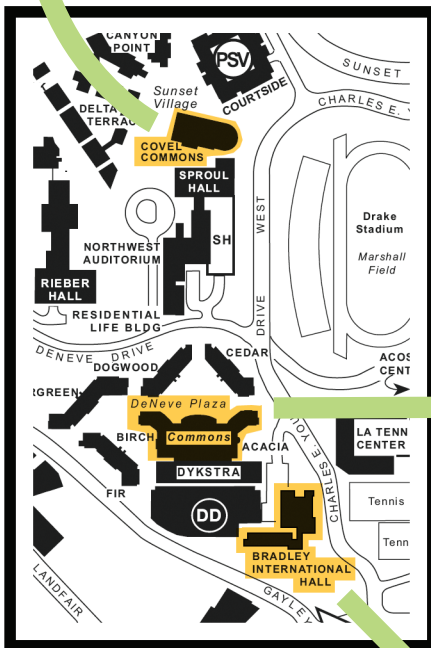
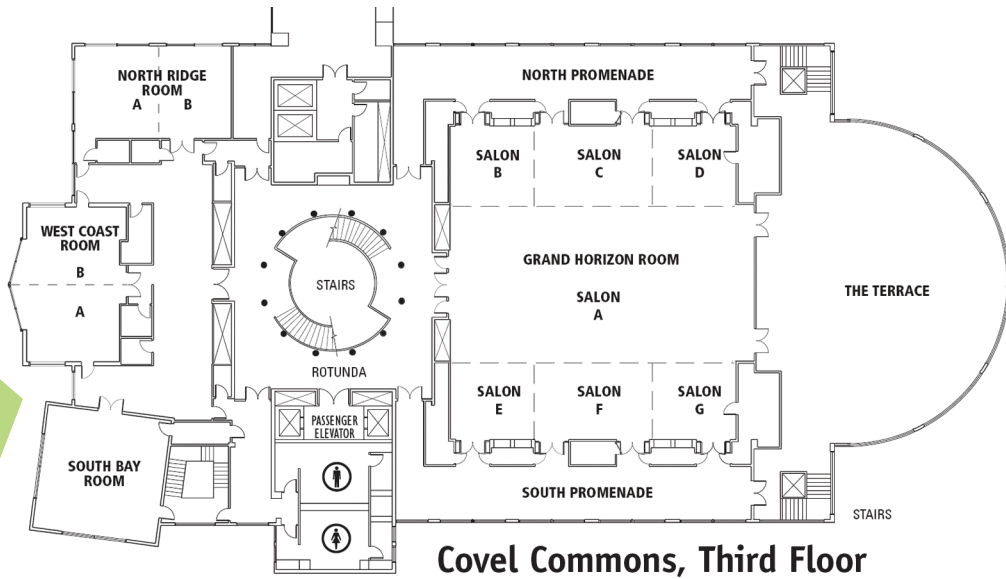
Session Chair: Sean A. Munson
Friday, 1:30 – 3:00 pm
Location: Covell, Grand Horizon Rm D

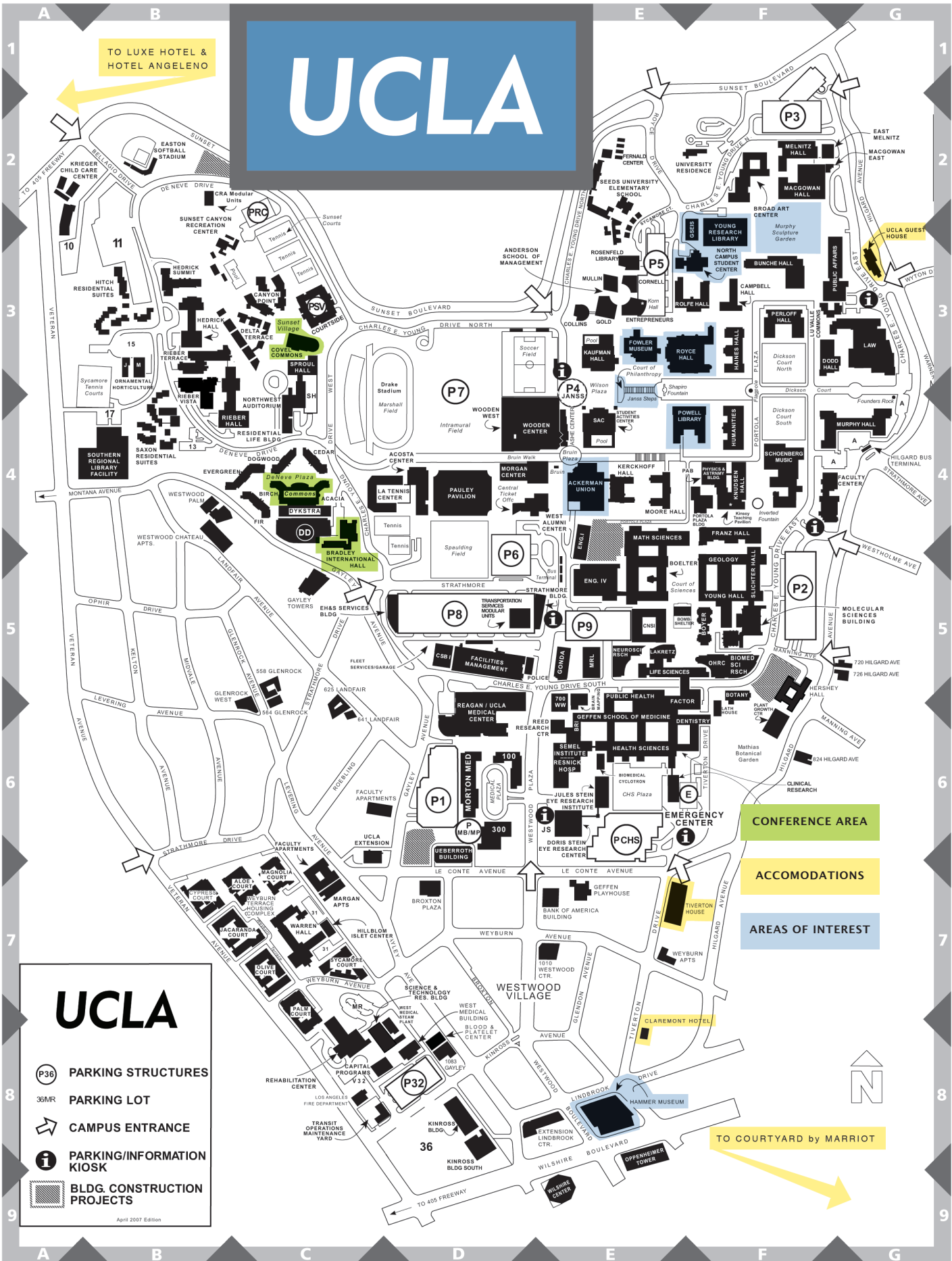
Training, Integration, and Identity: A Roundtable Discussion of Undergraduate and Professional Master's Programs in iSchools

Libby Hemphill and Sean A. Munson
University of Michigan; libbyh@umich.edu

Topics: nature and scope of iSchools and iResearch

Keywords: ischools, research, education, professional students





TO LUXE HOTEL & HOTEL ANGELENO

UCLA

UCLA GUEST HOUSE

CONFERENCE AREA

ACCOMODATIONS

AREAS OF INTEREST

TO COURTYARD by MARRIOTT

UCLA

- P36 PARKING STRUCTURES
- 36MR PARKING LOT
- CAMPUS ENTRANCE
- PARKING/INFORMATION KIOSK
- BLDG. CONSTRUCTION PROJECTS

April 2007 Edition

Los Angeles Area Map



UCLA Events During Your Stay

All events from happenings.ucla.edu

Tuesday February 26, 2008

Baseball-Home Games
vs. Cal State Northridge
Jackie Robinson Stadium

Center for the Study of Women Lecture
'Role Model for Somali Women of the Diaspora?'
Royce Hall Humanities Conf Rm 314

Exhibition :Portrait of a Jewish Artist: R.B. Kitaj in Text and Image
Art and writings by R.B. Kitaj
Charles E. Young Research Library

Exhibition: On the Wall - Silence = Death: Los Angeles AIDS Activism
Los Angeles AIDS Activism 1987-2007
Glorya Kaufman 2nd Floor South Lobby

Exhibitions: UCLA Hammer Museum
Fine art exhibitions
UCLA Hammer Museum

Jewish Studies Center Lecture
'Emerging Identities of Young Adult American Jews'
UCLA Hillel

Medieval & Renaissance Studies Sawyer Seminars
'Chaucers Parliament of Fowls'
Royce Hall H.Morris Rm 306

Professor in the Union Lunch Talk
Norton Wise, Department of History
Ackerman Union Meeting Room 2408

Wednesday February 27, 2008

Baseball-Home Games
vs. UC Santa Barbara
Jackie Robinson Stadium

Center for the Study of Women Screening
Ulrike Ottinger's film 'Ticket of No Return'
Melnitz Hall James Bridges Theater

Center for Women & Men Eating Workshop
'The Urge to Splurge:Controlling Overeating'
Student Activities Center Suite B44

Exhibition :Portrait of a Jewish Artist: R.B. Kitaj in Text and Image
Art and writings by R.B. Kitaj
Charles E. Young Research Library

Exhibition: On the Wall - Silence = Death: Los Angeles AIDS Activism
Los Angeles AIDS Activism 1987-2007
Gloria Kaufman 2nd Floor South Lobby

Exhibitions: Fowler Museum at UCLA
Art and material culture exhibitions
Fowler Museum at UCLA

Exhibitions: UCLA Hammer Museum
Fine art exhibitions
UCLA Hammer Museum

Film and Television Archive Screenings
'Army of Shadows'
UCLA Hammer Museum Billy Wilder Theater

Medieval and Renaissance Studies Faculty Roundtables
'The Devil's Interval'
Royce Hall Room 306 - Herbert Morris Rm

Men's Tennis-Home Matches
vs. UC Santa Barbara
Los Angeles Tennis Center

Planetarium Show and Telescope Viewing
Projected and magnified views of the night sky
Mathematical Sciences 8224 - Planetarium

Softball- Home Games
vs. Cal State Bakersfield
Easton Softball Stadium

Thursday February 28, 2008

Academic Advancement Program 35th Anniversary Celebration
Performances, awards and a retrospective
Royce Hall Theatre Area

Center for Women & Men Parents' Program
Parents' network
Student Activities Center Suite B44

Center for Women & Men Wellpower Workshop
Focusing Strategies
Student Activities Center Suite B44

Exhibition :Portrait of a Jewish Artist: R.B. Kitaj in Text and Image
Art and writings by R.B. Kitaj
Charles E. Young Research Library

Exhibition: On the Wall - Silence = Death: Los Angeles AIDS Activism
Los Angeles AIDS Activism 1987-2007
Gloria Kaufman 2nd Floor South Lobby

Exhibitions: Fowler Museum at UCLA
Art and material culture exhibitions
Fowler Museum at UCLA

Exhibitions: UCLA Hammer Museum
Fine art exhibitions
UCLA Hammer Museum

Fowler Out Loud Performances
Musical Diaspora
Fowler Museum at UCLA

In the Garden Noon Dance Performance
Performing: Baker & Tarpaga Dance Project
Glorya Kaufman Hall Garden Theater

Medieval & Renaissance Studies Distinguished Visiting Scholar Lecture
'Interpretations of the Quranic Verse'
Royce Hall H.Morris Rm 306

USAC Campus Events Commission \$2 Movie
'No Country for Old Men'
Ackerman Union Grand Ballroom (2400)

Women's Basketball Home Game
vs. Arizona State
Pauley Pavilion Arena

Friday February 29, 2008

Exhibition :Portrait of a Jewish Artist: R.B. Kitaj in Text and Image
Art and writings by R.B. Kitaj
Charles E. Young Research Library

Exhibition: On the Wall - Silence = Death: Los Angeles AIDS Activism
Los Angeles AIDS Activism 1987-2007
Glorya Kaufman 2nd Floor South Lobby

Exhibitions: Fowler Museum at UCLA
Art and material culture exhibitions
Fowler Museum at UCLA

Exhibitions: UCLA Hammer Museum
Fine art exhibitions
UCLA Hammer Museum

Medieval & Renaissance Studies Workshop
Medieval and Early Modern Slavic Studies
Royce Hall H.Morris Rm 306

USAC Campus Events Commission \$2 Movie
'No Country for Old Men'
Ackerman Union Grand Ballroom (2400)

Saturday March 01, 2008

Center for Women & Men Self-Defense Workshops
 Self-defense training
 To Be Determined

Exhibition :Portrait of a Jewish Artist: R.B. Kitaj in Text and Image
 Art and writings by R.B. Kitaj
 Charles E. Young Research Library

Exhibitions: Fowler Museum at UCLA
 Art and material culture exhibitions
 Fowler Museum at UCLA

Exhibitions: UCLA Hammer Museum
 Fine art exhibitions
 UCLA Hammer Museum

Historical Ballroom Dance in the Rotunda
 The Presidents' Ball
 Powell Building Rotunda

Medieval & Renaissance Studies Seminar
 Medieval history
 Overseer's Room, the Huntington Library

UCLA Extension Seminar: Soft Skills and Successful Relationships
 For fitness & health care professionals & students
 213 Extension Lindbrook Center

UCLA Extension Seminar: The Art of Mindfulness
 To enhance personal life, work, relationships
 Haines Hall 118

UCLA Extension Seminar: Writing a Book Proposal that Sells
 The 7 vital elements of the successful proposal
 Public Affairs Building 2270

Women's Basketball Home Game
 vs. Arizona
 Pauley Pavilion Arena

Sunday March 02, 2008

American Youth Symphony
Classical music from young artists
Royce Hall Theatre Area

Exhibitions: Fowler Museum at UCLA
Art and material culture exhibitions
Fowler Museum at UCLA

Exhibitions: UCLA Hammer Museum
Fine art exhibitions
UCLA Hammer Museum

Fowler Museum Lecture & Book Signing
Christopher B. Donnan
Fowler Museum at UCLA

Women's Gymnastics Home Meet
vs. Oregon State
Pauley Pavilion Arena

USAC Cultural Affairs Commission Eclectic Open Mic
Student Performances
Kerckhoff Hall Coffeehouse Stage

100 Places to Visit While You're in Los Angeles

<http://www.garyisse.com/article-224.htm>

Name	Category	City
Shambala Preserve and Roar Foundation	Attractions	Acton
Disneyland	Theme Parks	Anaheim
Knotts Berry Farm	Theme Parks	Anaheim
Wildlife Waystation	Attractions	Angeles National Forest
Santa Anita Race Track	Attractions	Arcadia
Arboretum of Los Angeles County	Gardens	Arcadia
Arcadia Historical Museum	Museum	Arcadia
Avalon - Catalina Island	Attractions	Avalon
The Wrigley Memorial and Botanical Gardens	Gardens	Avalon
Catalina Island Museum	Museum	Avalon
Academy of Motion Picture Arts & Sciences	Misc	Beverly Hills
Museum of Television and Radio	Museum	Beverly Hills
Hollywood Bowl	Arenas & Stadiums	Hollywood
Hollywood Wax Museum	Attractions	Hollywood
Ripley's... Believe it or Not Museum	Attractions	Hollywood
Max Factor Building	Landmarks	Hollywood
Capitol Records	Landmarks	Hollywood
The Silent Movie Theatre	Landmarks	Hollywood
ArcLight Cinerama Dome	Landmarks	Hollywood
Egyptian Theatre	Landmarks	Hollywood
El Capitan Theater	Landmarks	Hollywood
Grauman's Chinese Theatre	Landmarks	Hollywood
Hollywood Roosevelt Hotel	Landmarks	Hollywood
KCET Studio Tour	Misc	Hollywood
Universal Studios	Theme Parks	Hollywood
Great Western Forum	Arenas & Stadiums	Inglewood
Hollywood Park	Attractions	Inglewood
Randy's Donut	Landmarks	Inglewood
Aquarium of the Pacific	Attractions	Long Beach
Queen Mary	Attractions	Long Beach
Scorpion Russian Submarine	Attractions	Long Beach
Rancho Los Alamitos Historic Ranch and Gardens	Gardens	Long Beach
Long Beach Museum of Art	Museum	Long Beach
Museum of Latin American Art	Museum	Long Beach
Dodger Stadium	Arenas & Stadiums	Los Angeles
Walt Disney Concert Hall	Arenas & Stadiums	Los Angeles
Staples Center	Arenas & Stadiums	Los Angeles

Los Angeles Convention Center	Arenas & Stadiums	Los Angeles
Los Angeles Memorial Coliseum	Arenas & Stadiums	Los Angeles
Los Angeles Sports Arena	Arenas & Stadiums	Los Angeles
Shrine Auditorium	Arenas & Stadiums	Los Angeles
Skeletons in the Closet	Attractions	Los Angeles
Laserium at the Philharmonic	Attractions	Los Angeles
Griffith Observatory	Attractions	Los Angeles
Los Angeles Live Streamers	Attractions	Los Angeles
Los Angeles Zoo	Attractions	Los Angeles
La Brea Tar Pits	Attractions	Los Angeles
Farmers Market	Attractions	Los Angeles
San Antonio Winery	Attractions	Los Angeles
El Pueblo De Los Angeles State Historic Park	Attractions	Los Angeles
Old Chinatown Plaza	Attractions	Los Angeles
Olvera Street	Attractions	Los Angeles
Exposition Rose Garden	Gardens	Los Angeles
Kendle's Do-Nuts	Landmarks	Los Angeles
Coca-Cola Building	Landmarks	Los Angeles
Watts Towers	Landmarks	Los Angeles
Los Angeles City Hall	Landmarks	Los Angeles
Farmer John	Landmarks	Los Angeles
CBS Television City	Landmarks	Los Angeles
Union Station	Landmarks	Los Angeles
Angels Flight	Landmarks	Los Angeles
Los Angeles Public Library	Library	Los Angeles
California Science Center	Misc	Los Angeles
UCLA Hammer Museum of Art	Museum	Los Angeles
Getty	Museum	Los Angeles
MOCA @ Geffen	Museum	Los Angeles
Museum of African American Art	Museum	Los Angeles
Museum of Neon Art	Museum	Los Angeles
Los Angeles Craft and Folk Art Museum	Museum	Los Angeles
LACMA	Museum	Los Angeles
Los Angeles Museum of the Holocaust	Museum	Los Angeles
Natural History Museum of Los Angeles County	Museum	Los Angeles
Getty Malibu	Museum	Malibu
Roundhouse Aquarium	Attractions	Manhattan Beach
Manhattan Beach	Surf/Beach	Manhattan Beach
Marina Del Rey	Surf/Beach	Marina Del Rey
Fisherman's Village	Attractions	Marina Del Rey
Eaton Canyon	Attractions	Pasadena
Rose Bowl	Arenas & Stadiums	Pasadena
Pasadena Civic Auditorium	Arenas & Stadiums	Pasadena
Old Town Pasadena	Landmarks	Pasadena
Norton Simon	Museum	Pasadena

Pacific Asia Museum	Museum	Pasadena
Pasadena Historical Museum	Museum	Pasadena
Pasadena Museum of California Art	Museum	Pasadena
Raging Waters	Theme Parks	San Dimas
Huntington Library	Museum/Gardens	San Marino
Santa Monica Civic Auditorium	Arenas & Stadiums	Santa Monica
Santa Monica Pier Aquarium	Attractions	Santa Monica
John Drescher Planetarium	Attractions	Santa Monica
Third Street Promenade	Attractions	Santa Monica
Santa Monica Museum of Art	Museum	Santa Monica
Hurricane Harbor	Theme Parks	Valencia
Six Flags Magic Mountain	Theme Parks	Valencia
Venice Beach	Attractions	Venice
Rocky and Bullwinkle	Landmarks	West Hollywood
MOCA @ the Pacific Design Center	Museum	West Hollywood
The Westwood Bldg	Landmarks	Westwood
UCLA Fowler Museum of Cultural History	Museum	Westwood
Mount Wilson Observatory	Attractions	Mt. Wilson

Places to Eat in Westwood Village (310 Area Code for all Numbers)

COFFEE AND BREAKFAST

- **Border's Cafe Espresso**
1360 Westwood Blvd. 475-3444.
- **Arrosto Coffee**
923 Broxton Avenue. 824-2277.
- **City Bean Coffee**
10911 Lindbrook Drive. 208-0108.
- **(The) Coffee Bean & Tea Leaf**
Family-run specialty company selling coffee beans, rare teas, and accessories.
1001 Gayley Avenue. 208-1991.
950 Westwood Blvd. 208-8018.
- **Headlines**
10922 Kinross Avenue. 208-2424.
50's-style diner. Breakfast, sandwiches, malts.

- **Stan's Corner Donut Shoppe**
10948 Weyburn Avenue. 208-8660.
Donuts & breakfast sandwiches.

For almost 40 years, Stan's Corner Doughnut Shop of Westwood, California, has baked the world's most gourmet donuts. Every customer finds something to satisfy their donut desires--from a Cinnamon Chocolate Cheese Danish, to an Apple Fritter, to a Reese's Peanut Butter Pocket with fresh banana. Stan's offers donuts you will find nowhere else in the world.

- **Starbucks**
Starbucks Coffee Company is the leading retailer, roaster and brand of specialty coffee in the world.

1161 Westwood Blvd. 208-5505.

AMERICAN

- **Breadstiks**
1057 Gayley Avenue. 209-1111.
Groceries, sandwiches and salads.
- **Burger King**
900 Westwood Blvd. 208-6781
- **Butterfly Bakery**
Sandwiches and salads.

The cookies are irresistible. Not in the mood for something sweet? It also has sandwiches and salads for the lunch crowd, and there are frequent daily specials.

1269 Westwood Blvd. 479-6902.

- **Cuccini's Cafe**
960 Gayley Ave. 209-5591.
coffee & juices.
- **Elysee Bakery**
1099 Gayley Avenue. 208-6505.
Croissant sandwiches, soup, quiche.
- **Fatburger**
10955 Kinross Avenue. 208-7365.
Hamburgers, chicken wings, malts.
- **Gypsy Café**
940 Broxton Avenue. 824-2119.
Croissant sandwiches, quiche, pasta salads, pastries.
- **In'N Out**
At In N Out we use only the freshest ingredients, delivered to our stores every day. Never frozen, and no burger hits the grill until you order it.

922 Gayley Ave. (800) 786-1000.
Hamburgers, malts.
- **Jerry's Famous Deli**
10925 Weyburn Ave. 208-3354.
Wide deli selection, especially sandwiches.
- **Madison's Bar & Grill**
1037 Broxton Ave. 824-6250.
Entrees, sandwiches, salads. Bar.
- **Maloneys on Campus**
1000 Gayley Avenue. 208-1942.
Hamburgers, sandwiches, salads. Bar.
- **Paris Pastry Bakery**
1448 Westwood Blvd. 474-8888.
- **Sandbag's**
1134 Westwood Blvd. 208-1133.
Sandwiches and salads.

Places to Eat in Westwood Village (310 Area Code for all Numbers)

- **Schlotzsky's Deli**
1061 Broxton Avenue. 824-6375.
Deli sandwiches, salads, pizza.
- **Sepi's**
10968 Le Conte Avenue. 208-7171.
East Coast-style submarine sandwiches.
- **Subway**
1151 Westwood Blvd. 208-7774.
Submarine sandwiches, salads, croissant sandwiches.
- **Togo's**
10889 Lindbrook Drive. 208-4416.
Sandwiches.
- **Tommy's of Westwood**
970 Gayley Avenue. 824-4114.
Chili burgers.
- **Village Yogurt**
10887 Weyburn Ave. 209-2913.
- **Westwood Brewing Co.**
1097 Glendon Avenue. 209-BREW.
Sandwiches, salads, pizza. Bar.
- **Mongolian BBQ**
Mongolian BBQ, stir-fried vegetables and meat.
- **New India Grill**
1321 Westwood Blvd. 996-0064.
Indian cuisine.
- **Noodle Planet**
1118 Westwood Blvd. 208-0777.
Thai cuisine.
- **Sak's Teriyaki**
1121 Glendon Avenue. 208-2002.
Japanese fast food. Teriyaki chicken and beef.
- **Scallion's**
1091 Broxton Avenue. 824-7869.
Chinese food including dim sum.
- **Spike's Teriyaki Bowl**
1133 Westwood Blvd. 209-4952.
Japanese cuisine.
- **Tengu**
10853 Lindbrook Dr. 209-0071
Japanese cuisine.

ASIAN

- **Asuka**
1266 Westwood Blvd. 474-7412.
Japanese cuisine.
- **Cowboy Sushi**
911 Broxton Avenue. 208-7781.
Japanese cuisine, including teriyaki, tempura, and sushi.
- **Eurochow**
1099 Westwood Blvd. 209-0066
- **First Szechuan Wok**
10855 Lindbrook Dr. 208-7785. Chinese food, lunch speci
- **Isshin**
10861 Lindbrook Drive. 208-5224.
Japanese cuisine, including teriyaki, tempura and sushi.
- **Mongols**
1064 Gayley Avenue. 824-3377.

CONTINENTAL

- **The Gardens on Glendon**
1139 Glendon Avenue. 824-1818.
Full menu Steak, chicken, pastas, fresh fish. Bar.
- **Westwood Marquis Hotel Garden Terrace**
930 Hilgard Avenue. 208-8765.
Elaborate buffet and regular menu weekdays. Pricy. Bar.
- **Moustache Café**
1071 Glendon Avenue. 208-6633.
Pasta, seafood, salads, omelets.

ITALIAN/GREEK

Places to Eat in Westwood Village (310 Area Code for all Numbers)

Italian cuisine. Bar.

- **B.J.'s Pizza & Grill**
939 Broxton Ave. 209-7475.
Pizza, Italian sandwiches, salads.
- **Delphi**
1383 Westwood Blvd. 478-2900.
Greek cuisine.
- **California Pizza Kitchen**
1001 Broxton Ave. 209-9197
Pizza, pasta, salads.
- **Don Antonio's Pizzeria**
1136 Westwood Blvd. 209-1421.
Pizzas.
- **Di Stefano's Ristorante**
1076 Gayley Avenue. 208-5117.
Wide variety of authentic Italian food. Homemade pasta.
- **Italian Express**
10845 Lindbrook Drive 208-5572.
- **Lamonica's**
1066 Gayley Avenue. 208-8671.
New York-style pizza sold by the slice.
- **Olive Garden**
936 Westwood Blvd. 824-7588.
Wide variety of Italian cuisine.
- **Pomodoro**
1393 Westwood Blvd. 445-9998.
Pizza, pasta, Italian cuisine.
- **Palomino**
10877 Wilshire Blvd (@ Glendon & Lindbrook)
(310) 208-1960
- **Shakey's**
1114 Gayley Avenue. 824-4111.
Pizza.
- **Sprazzo**
1389 Westwood Blvd. 479-3337.
Italian cuisine.
- **Tanino**
1043 Westwood Blvd. 208-0444.

MEXICAN

- **Acapulco**
1109 Glendon Avenue. 208-3884.
Mexican food. Seafood specialties. Bar.
- **Baja Fresh Mexican Grill**
10916 Lindbrook Drive. 208-3317.
Tacos, tostadas, fajitas, burritos.
- **El Pollo Loco**
1081 Gayley Avenue. 824-7776.
Grilled chicken.
- **José Bernstein's**
935 Broxton Avenue. 208-4992.
Tacos, burritos, deli sandwiches.
- **La Salsa**
1154 Westwood Blvd. 208-7083.
Tacos, burritos.
- **Me Gusta**
1055 Broxton Avenue. 208-2226.
Mexican cuisine.

EASTERN

- **Falafel King**
1059 Broxton Avenue. 208-5782.
Falafel, shwarma, salads.

It can be described as L.A.'s fast food version of Middle Eastern cuisine. It specializes in falafel, of course, which is a deep-fried chickpea mixture served in steaming hot pitas with tahini sauce, lettuce and tomato.
- **Shahrezad**
1442 Westwood Blvd. 470-3242.
Middle-Eastern cuisine.

VEGETARIAN

Places to Eat in Westwood Village (310 Area Code for all Numbers)

- **California Tea House**
10874 Kinross Avenue. 824-4284.
Sandwiches, teas and juices.
- **Jamba Juice**
10912 Lindbrook Ave. 209-1778.
Sandwiches, fresh-squeezed juices.
- **Native Foods**
1110 ½ Gayley Ave. 209-1055.
Vegetarian cuisine.

Native Foods is an organic vegan / vegetarian restaurant designed to meet the needs of today's discriminating, health-conscious diners – vegetarians and meat-eaters alike.

- **Nature's Health Cove**
1010 Broxton Avenue. 208-7333.
Salads, sandwiches, juices, vitamins, health products.

ICE CREAM AND SWEETS

- **Baskin-Robbins Ice Cream Store**
10916 Kinross Avenue. 208-8048.
- **Candy Castle**
945 Westwood Blvd. 208-0293.
- **Diddy Riese**
926 Broxton Avenue. 208-0448.
Cookies, muffins, ice cream.

Diddy Riese Cookies first opened it's doors in 1983 and has been serving high quality cookies, brownies and ice cream—at very reasonable prices—to the Los Angeles community ever since

- **Haagen Dazs Ice Cream Shop**
10878 Kinross Avenue. 208-7405.
- **Mrs. Fields Cookies**
907 Westwood Blvd. 208-0096.

MEET THE PARTICIPANTS ... STUDENTS' PERSONAL STATEMENTS

Benjamin Kwasi Addom

School of Information Studies

Syracuse University

bkaddom@syr.edu



Poster title:

"Community innovativeness": A new perspective on knowledge creation

Personal statement:

The theory of "absorptive capacity" is well recognized in the fields of Information Systems and Organization Studies. It has been extensively applied and tested to organizational, inter-firms, and national situations. However little is known about its implications on "communities" for knowledge generation and subsequent innovativeness. My dissertation interest is in the exploration of this theory in geospatial communities to determine how social structure of a given community can predict its absorptive capacity and thereby leading to its ability to innovate. The understanding of innovative capabilities of physical communities, in which we work, will definitely influence our design of technological projects for these communities.

Ben Addom received a combined Bachelors Degree in General Agriculture and Diploma in Education from the University of Cape Coast, Ghana between 1993/1998. As a Ghanaian, he served his nation as Agricultural Extension Officer for one year through a mandatory National Service Program that applies to all students graduating from the country's public higher education institutions. He also taught agriculture in Secondary School (High School) for few months before returning to the field where he worked with two different private organizations between early 2000 and 2003 prior to his admission to Cornell University. Ben was responsible for communicating research outcomes and innovations from organizations and research stations to communities for their agricultural productions and on community development issues with feedbacks from the communities to the organizations and research stations.

While at Cornell, Ben enrolled in Masters in International Agriculture and Rural Development in 2003 and graduated in 2004. He conducted a survey on the E-readiness of seven universities in Ghana as part of his Masters research to assess how prepared the country's higher education institutions are in using the new Information Communication Technologies (ICTs) for teaching, research and outreach/extension, with very interesting and promising results. He also worked as a student computer assistant at Stone Computing Center; digital preservation assistant at Collection and Development unit; and administrative assistant at the Circulation unit all at Mann Library, Cornell University. His interest in the use of ICTs emerged while doing his Masters studies at Cornell University. However, based on his experience as an extension officer in deprived communities of Ghana, Ben is careful with the idea of knowledge/technology/innovation transfer and diffusion from the developed to developing countries.

At the iSchool at Syracuse University, Ben is exploring different ways by which these new technologies could

be used to enhance access by all. Specifically, Ben has been looking at issues relating to transnational civil society, epistemic communities, geographically distributed collaboration, knowledge sharing and information exchange between developed and developing nations, timely access to relevant and content specific information by end users, technology and community networking, and social network analysis. He is exploring these ideas through the fields of ICT4D and Community Informatics and Networking.

Ben has been the Lab Manager of the Center for Research on Collaboratories and Enhanced Learning Communities (Cotelco) (www.cotelco.syr.edu) since 2005 and a research fellow at the Burton Blatt Institute, Syracuse University (<http://bbi.syr.edu>). He looks forward to be involved in grassroots policy, research, and teaching issues that impact the design and use of ICTs for underserved communities both in developed and developing nations after his graduate program.

Johanna L. H. Birkland

School of Information Studies

Syracuse University

jlbirkla@syr.edu



Poster title:

Older adults and information technology: The current state of research and future directions

Personal statement:

Johanna Birkland is currently a doctoral student at the School of Information Studies, Syracuse University. She holds two previous graduate degrees, a M.S. in Communication Management from Ithaca College and a M.S. in Instructional Design from Syracuse University. Previously she worked as a preservation assistant at the Albert R. Mann Library, Cornell University, where she was responsible for managing the indexing process of two historical digitalization projects. She also has consulted for local non-profit organizations in the areas of instructional design and communication.

Johanna's research interests broadly include older adults and technology. Her interests in this area include the application of technology and technological training for older adults in the workplace. She is also interested in social issues regarding technological use that impact older adults, such as the digital divide. Johanna is also interested in exploring generational shifts, and how these impact the perceptions of technology and its use by individuals of all ages.

Miao Chen

School of Information Studies

Syracuse University

mchen14@syr.edu



Poster title:

Deriving ontology from folksonomy and controlled vocabulary

Personal statement:

I am currently a PhD student at School of Information Studies, Syracuse University. Starting from fall 2005, this is my third year in the program. My advisor is Prof. Jian Qin. I got my bachelor degree in library and information science from Peking University, China, in 2005. My research interest includes ontology, information organization, ontology, and related areas. More specifically, what I am studying now is theories and methods of building ontology and techniques of extracting concepts and relations from data set.

Coming from a field composed of both traditional LIS knowledge and new technologies, I am aware of the critical influence of traditional LIS on the present information science; meanwhile, my academic training in the previous and current programs provides me a view of the important role of technology which facilitates the fast development of the field. In ontology-related research, classification theories and practices in library science serve partially as the cognitive and behavioral foundations of ontology, techniques such as machine learning and data mining help us explore components in ontology, and theoretical and methodological research of ontology absorbs knowledge from the previous two fields and develops its own content. My research is actually oriented by this thread, combining traditional classification and automatic techniques to contribute to ontology research as well as learning existing knowledge in the ontology studies.

I have been exploring more details and experimenting in automatically building ontology from controlled vocabulary and folksonomy. It aims to combine merits of the two types of vocabulary sources to build a comprehensive ontology. The general methodology is proposed in this i-conference poster, while there is much more to develop, such as relation extraction and ontology mapping, to actually implement the methodology. When studying this topic, relation extraction in ontology intrigues me greatly and I want to do more research on it. Though having not decided my dissertation topic, this one seems really interesting and might serve as a candidate topic.

Another project I am working on is evaluating an IR system developed by the Center for Natural Language Processing, Syracuse University. Besides ontology and IR related projects, I have been participating projects of various disciplines and I believe I benefit from looking at research from different fields of this interdisciplinary information science. I have been involved in Wireless Grid and Study Response projects in our school and experienced research methods and tools of other information fields. I am also assisting one applied research project of JPMC-SU partnership, on collaboration topic. Going back to my own research fields from projects in other fields, I can find how these fields are interrelated and acquire a better understanding of my research area.

As a third year PhD student, I am at the time of finding my dissertation topic and publicize my academic efforts such as papers and posters. I will probably study comprehensively on methodology of ontology building and influence of ontology on information retrieval for the dissertation. As a useful approach of organizing knowledge, ontology research has large potential to be applied to individuals and organizations and I am excited to be a learner and researcher in this area.

My career plan has been in my consideration since I entered the PhD program. I plan to be a researcher in a research institute or an expert in a public service department. I hope to continue my research on ontology to organize knowledge and facilitate information retrieval in my future job with my expertise. It is always nice to get things organized and approached.

Mary Jo Dorsey

School of Information Sciences

University of Pittsburgh

mjdorsey1@gmail.com



Poster title:

A preliminary consumer health information-seeking (CHI-seeking) behavior model of physicians who treat elderly depressed patients (results from the pilot study of a dissertation proposal)

Personal statement:

Education

PhD Candidate, University of Pittsburgh, School of Information Sciences

(April 2008 expected graduation)

M.L.S. University of Pittsburgh, School of Information Sciences 1990

B.S. Carlow College (Biology)

1988

Positions and Employment

1990-1997	Technical Information Center Manager TRW Environmental Safety Systems, Inc., and Reynolds Electrical & Engineering Co., Inc., for U.S. Department of Energy, Las Vegas, NV, Yucca Mountain Site Characterization Project
1997-2001	Assistant Curator, Coordinator of Information Services, Head, Computer Media Center, Ehrman Medical Library, New York University School of Medicine
2001-2002	Assistant University Librarian, Information Services, Department Cornell Medical Library, Weill Medical College of Cornell University, New York, NY
2003-2004	Researcher Alexander Foundation for Women's Health, Berkeley, CA
2002-2008	Director, Medical Library, The Western Pennsylvania Hospital, Richard M. Johnston Health Sciences Library and Patient and Family Education Center, Pittsburgh, PA
2005-present	Adjunct Faculty Member, Department of Library and Information Sciences, School of Information Science, University of Pittsburgh
2006-2007	Graduate Student Researcher (two-year appointment), University of Pittsburgh School of Medicine, Department of Biomedical Informatics. Grant R21 LM 008565, National Library of Medicine, Valerie Monaco (PI). Reading Level Evaluation Engine for E-Health Resources

2008

Visiting Reference Librarian, Health Sciences Library System, University of Pittsburgh, Pittsburgh, PA, liaison to School of Information Sciences

Research

1. *Modeling the consumer health information-seeking (CHI-seeking) behavior of physicians who treat elderly depressed patients*, current dissertation study. Expected defense: April 2008.
2. *A preliminary consumer health information-seeking (CHI-seeking) behavior model of physicians who treat elderly depressed patients (results from the pilot study of a dissertation proposal)* Presentation: University of Pittsburgh, Institute on Aging Fall Poster Session, December 2007.
3. *Comparison of Six Medical Search Engines*, paper in-progress with Valerie Monaco, PhD, MHCI, University of Pittsburgh School of Medicine, Department of Biomedical Informatics, Grant R21 LM 008565, National Library of Medicine.
4. *Before the Search: Characterizing the Health Concerns and Questions of Low-literacy Individuals* Project co-investigator with Valerie Monaco (PI), PhD, MHCI, University of Pittsburgh School of Medicine, Department of Biomedical Informatics. Grant R21 LM 008565, National Library of Medicine. Paper presented at the annual conference of the Medical Library Association, Philadelphia, PA. May 21, 2007.
5. *Web (Non)Sense: Is LGBT Health Information MLA from the Internet at Your Local Hospital?* Presented at the annual conference of the Medical Library Association, Phoenix, AZ, May 2006.

Selected peer-reviewed publications (in chronological order)

1. Dorsey, MJ. Nuclear Science. In: Tillman, H., ed. *Internet Tools of the Profession, A Guide for Special Librarians*, Washington, DC, Special Libraries Association, 1995: 127-134.
2. Dorsey, MJ. Message from a New Yorker, Reader Forum, *American Libraries*, November 2001.
3. Dorsey MJ, Detlefsen E. "Investigating Information-seeking Behaviors of Primary Care Physicians Who Care for Older Depressed Patients and Their Family Caregivers: A Pilot Study." *Journal of the Canadian Health Libraries Association* 26:111-116 (Fall 2005).
4. Dorsey, MJ, "Searching the Literature: An Evidence-Based Approach to Clinical Practice." *AOHP J.* Fall 2006; 26(4): 13-15.

Interests

Travel, drama and musical theatre, early music, reading, my family

Career goals

I am pursuing a career as an academic researcher and professor at a graduate school of information sciences or graduate school of public health or medicine.

In the course of my academic and professional work I have become drawn to studying the behaviors of clinical and consumer health information seekers. The value of health information from the web is affected by two factors: (1) the disparities of how information is organized and authenticated on the web and (2) by users' innate problem-solving approaches and searching habits.

My goal is to accurately describe what differences there may be when the domain is different and contribute to the improvement of the outcomes of both clinical and consumer searches by describing the model of what is happening now. A long term goal is to look at subspecialties of medicine and the needs of varying user groups such as the growing senior population and gays and lesbians.

Megan Finn

School of Information
University of California, Berkeley
megfinn@ischool.berkeley.edu



Poster title:

Post-disaster information infrastructure: The 1989 Loma Prieta earthquake

Personal statement:

Government reports and members of the media blame information systems as the culprit of what has been characterized as faulty responses to disasters such as Virginia Tech and Hurricane Katrina. In the process they propose information and communication technologies (ICTs) as solutions to disaster response. Calls for new ICTs following recent American disasters reveal a poor understanding of the socially situated nature of ICTs. Disaster-related research does not expound a sophisticated understanding of ICTs either. As a consequence, information systems are seen as an unproblematic means of informing the right people with the right information at the right time.

In my dissertation, I will challenge these simplistic notions of information systems as solutions to disaster response by proposing two case studies of “information environments” from the period following the 1989 Loma Prieta earthquake. I will introduce an analytical framework from the social studies of information which situates information systems in the social environment. The theoretical construct, the “information environment,” is meant to overcome the limitations of studying ICTs from a purely technical perspective.

The information environment is a synthesis of theoretical work that enables me to look beyond the simple view of ICTs as mere conveyors of information to ask how ICT use is shaped by and shapes the social world. Though the focus of my research is “postdisaster,” the information environment construct emphasizes a longer time horizon. For example, one of the information environment approaches poses questions about the *values* that may be embedded in the design of an ICT. Another approach asks about the *conventions of practice* around an information infrastructure. Both “values” and “conventions of practice” point to activities which occurred prior to a disaster, calling to mind the question of how far in the past one must look. Additionally, one or both of the “values” or “conventions of practice” concepts might be useful to explain actions following a disaster. The information environment is an intentionally broad (but by no means comprehensive) construct able to include many approaches which socially contextualize ICT use. Part of this project is then determining which aspects of the information environment are most important to understanding ICT use following a disaster. By iteratively going back and forth between the numerous approaches included in the information environment and the case data, I will expand and refine the salient aspects of the information environment construct for the “post-disaster” cases. Thus, this dissertation will not only apply ideas from the social studies of information to expand disaster related research, but also the somewhat novel post-disaster environment will challenge approaches from the social studies of information.

My primary research question is:

- (a) How do people use ICTs to make sense of the world in the period following a disaster?

The information environment construct also proposes the questions:

- (b) How is ICT use shaped by formal and informal social structures?

My case analysis should also help to answer the question:

- (c) How can the information environment be refined and synthesized to best understand ICT use following a disaster?

If I can answer questions (b) and (c), I will gain an understanding of ICT's situated in the local social context. This knowledge will then help to answer my primary question about the cases.

Using the case studies, this dissertation will facilitate an analysis of two propositions:

- (i) Understanding the information environment is crucial to discussions about information systems for disaster response.
- (ii) Variations in how people use ICT's helps to explain how particular communities can experience the same disaster differently.

In order to address my research questions I propose a theoretically informed, multi-method, pragmatic examination of historical data about the 1989 Loma Prieta earthquake. Using the information environment framework, I will closely examine two cases of community use of an information and communication technology post-disaster. I will look at the following two cases studies: Radio use by the Spanish-speaking community, particularly in Watsonville; and use of the internet, specifically the online community on The Well. Secondary survey data will serve to contextualize my cases within the greater Loma Prieta disaster. By analyzing post-disaster ICT-related practices in two distinct communities, I will show how the information environment framework both elucidates these practices, as well as explains differential post-disaster experiences.

The unique combination of historical work on information environments in a post-disaster social context will allow me to make contributions to the social studies of information, and to research on disasters. Finally, my work will provide a case study for doing historical research utilizing new media tools and techniques.

Saira N. Haque

School of Information Studies

Syracuse University

snhaque@syr.edu



Roundtable title:

The role of iSchools in medical informatics

Personal statement:

I have been in the doctoral program at Syracuse University's School of Information Studies since 2004. Prior to joining the program, I held a variety of leadership positions in healthcare administration and management consulting. I am interested in coordinating work in interdisciplinary teams, organizational impacts of information systems and privacy and security. My dissertation is a mixed-method study where I look at how interdisciplinary healthcare teams coordinate their work in their daily routines. I enjoy both research and teaching and am looking forward to completing my dissertation and starting a faculty position next year.

Interdisciplinary teams that are separated by time, space and profession have unique coordination challenges. Sometimes, mechanisms such as information systems can help ameliorate them. Other times, these are not integrated into the daily workflow and routines of team members and have unintended consequences. In these cases, coordination errors or failures may result. Studying this involves understanding organizational

environment and everyday routines and work practices. This knowledge can help when using information technology to support work and prevent coordination failures. One place where this can be seen is in the healthcare industry, which is characterized by hierarchy, shiftwork, multiple departments and various organizations.

Healthcare delivery takes place through distributed teams whose members are separated by organizational, temporal, spatial and professional boundaries. In order to understand how these caregivers work together, I draw upon literature about routines, boundaries and coordination. By looking at these areas in tandem rather than separately, we can have a richer understanding of how work occurs in healthcare organizations and how to improve it. Routines are recurrent interaction patterns. These help in understanding the context in which work occurs. In addition, studying routines helps to distinguish between the purported routine and the routine in practice. If policies, procedures and coordination mechanisms such as information systems are based on the purported routines, they might not support the actual performances of the routines. Thus, it is important to understand them at both levels.

These patterns involve people working across functional, spatial, temporal, and organizational boundaries. These boundaries might have variable impacts on the routines and how to coordinate the work within them. Thus, they result in coordination challenges. In this case, coordination refers to managing the dependencies between the interrelated activities of the routine. Understanding the activities, their dependencies and the boundaries that must be crossed can help when developing information systems and other mechanisms to address the needs of workers within the routine.

As documentation in the US healthcare system moves from paper-based to electronic, we will have a period where documentation occurs in both media. Thus, the integration of coordination mechanisms with workflow will become particularly important. This change will also bring about privacy and security concerns. I am interested in understanding how interdisciplinary healthcare teams coordinate their work so that we can develop better coordination mechanisms and design information systems to facilitate that coordination. This could lead to better outcomes and improved patient safety.

While electronic medical records are commonly cited as the way to improve coordination, if they are not implemented thoughtfully, then they might not achieve the desired objectives. This could be because of a lack of awareness of how records and the information within them are used in healthcare work. Those who work together to care for patients use the medical record as both a communication device and an information repository. Often, the medical record is some amalgamation of paper and electronic documentation that is tailored to the needs of individual team members. The variety of media, users and organizations result in unique privacy and security concerns. Certainly the improved availability of information can improve outcomes and lead to better decision-making. But, it is important to ensure that organizations have systems and processes in place to ensure that the information is only in the hands of those who need it.

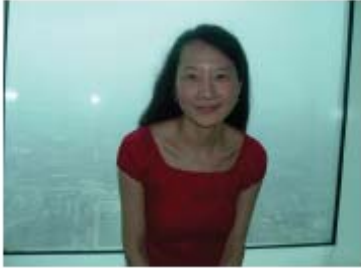
By improving coordination in distributed interdisciplinary teams, we can prevent coordination errors which can lead to important tasks not being completed or duplication of work. In addition, this is useful when developing coordination mechanisms such as information systems to improve outcomes. Because these issues are widespread in distributed work, I expect that my research would have impact in domains outside of healthcare as well. In order to research to be applied, not only does the research have to be disseminated, but those in the field need to understand it. To that end, I feel a strong sense of responsibility to make my work accessible to both researchers and practitioners. This will allow me to make a contribution to both the academic and practical realms.

Shuyuan Mary Ho

School of Information Studies

Syracuse University

smho@syr.edu



Poster title:

Behavioral parameters of trustworthiness for countering insider threats

Personal statement:

Interdisciplinary Research Identity: Information Systems Security (ISS)

My research interests focus on information systems security (ISS), which includes information content security, information security management and computing security. *Computing security* comes from research in the information systems discipline, and information science discipline. *Computing security* emphasizes how to model a secure information system that provides confidentiality, integrity and availability for the information, and knowledge management for corporate assets. *Information content security* includes research in information science & technology, organizational science and socio-psychology. *Information content security* emphasizes control over information content through information architecture, information classification, with an understanding of subtle human language in detecting peculiar intention and motivation in the workplace. *Information security management* is drawn from research in the information systems discipline, organizational science discipline and socio-psychology. *Information security management* examines management best practices that contribute to a secure environment, prevent security incidents resulting from social-engineering techniques by outsiders and unauthorized access by insiders.

Teaching Philosophy

My teaching philosophy can be best characterized as a blend of theory and practice, using lectures and labs in conjunction with individual mentoring, and including periodic quizzes as a prelude to each significant test. I believe that teaching should revolve around providing both guidance and consistent feedback loops, as well as the freedom to discover new concepts - whether face-to-face or in a virtual classroom. Teaching helps to transmit knowledge and skills to younger generations. Research sets up a foundation for advanced study, so that knowledge given to students is not only historical, but progressive - and helps students to migrate into their future.

I believe that research should be built into the foundation of advanced teaching. Disciplined knowledge and theory are communicated to the students through teaching, but it is through *research* that increasing knowledge of the discipline can be best established and applied. Acquiring professional expertise requires a mindset of life-long learning. New ideas, new strategies, and newly applied behaviors - all are part of the infusion of research into any curriculum for bringing undergraduate level students into the IT workforce. Knowledge must be applied to develop skills. It is my belief and practice to transform students into becoming information professionals, or even information systems security professionals, which benefits individuals as well as the IT community.

Academic and Professional Competence in Information Systems Security (ISS)

In addition to my ISS research and teaching experience at iSchool Syracuse University, I am a Certified Information Systems Security Professional (certified by (ISC)2), and a Certified Information Security Manager (certified by ISACA). I have been an ISS Professional actively involved in technologies such as IPSec VPN, SSL VPN, Certificate Authority, RBAC, firewall, intrusion detection, vulnerability assessment, and policy management since 1996. I successfully managed a software-based virtual private network technology transfer project for the Republic of China Military Defense during 1999-2000 and led a research team in developing an event correlation platform; SIMPL/E (Security Information Monitoring Platform/Environment) in 2003. I also participated in a research project, Countering Insider Threats, which was supported by the Information Assurance for the Intelligence Community (IAIC) program of the Advanced Research and Development

Activity (ARDA). This research has been published and presented at the 2nd Symposium Intelligence & Security Informatics, Tucson AZ, June 2004. I was nominated as one of the 2004-2006 (ISC)² Scholarship Award Winners, and was recognized at the 2005 RSA Conference, San Francisco, CA.

I oversaw the distribution of software throughout Greater China from 1999 through 2004. This software was policy-based information security software which included VPN, firewall, CA (certificate authority) and RBAC (role-based access control) capabilities. During my responsibilities as a System Engineering Manager at RedCreek Communications and North Asia Technical Director at Internet Dynamics from 1999 through 2001, I supported market expansion, software distribution to greater China, and assisted over 20 VPN (virtual private network) installations in Asia Pacific regions including China, Hong Kong, Macaw, Singapore, Korea and Taiwan. I provided sales/engineering training, pre-sales support, system demonstration and lab tests. I also assisted in developing technology resources and strategic alliances with companies in mainland China. During the market expansion, I helped in designing customer network architecture, integrating security operations components, and assessing vulnerability in technical services provided.

Dissertation Proposal Abstract: Behavioral Anomaly Detection – Perceived Trustworthiness in an Online Community

My dissertation study examines perceptions of human *trustworthiness* as a key component for countering insider threats. The term *insider threat* refers to situations where a critical member of an organization behaves against the interests of the organization, in an illegal and/or unethical manner. Identifying and detecting how an individual's behavior varies over time - and how anomalous behavior can be detected - are important elements in the preventive control of insider threat behaviors. The proposed research focuses on understanding how anomalous behavior can be detected by observers. While human observations are fallible, this study adopts the concept of human-observed changes in behavior as analogous to "sensors" on a computer network. Using online team-based game-playing, this study seeks to re-create realistic situations in which human sensors have the opportunity to observe changes in the behavior of a focal individual – in this case a team leader. Transcripts of communications during the game-playing situation are examined in order to understand how human sensors attribute meaning to an individual's potentially suspicious behavior. Results of this study may lead to the development of semi-automated or fully-automated behavioral detection systems that attempt to predict the occurrence of malfeasance.

I-Han (Sharon) Hsiao

School of Information Sciences

University of Pittsburgh

ihh4@pitt.edu



Poster title:

Cyber Java Monopoly: Game-based approach of collaborative programming language learning

Personal statement:

My name is I-Han HSIAO, just call me Sharon. I'm originally from Taiwan. Now, I'm a second year PhD student from School of Information Sciences, University of Pittsburgh. I got my MSc in Business Information Systems from University of London. I used to dabble in JavaCard Technology for 2 years or so. I worked as SAP ERP engineer for almost 2 years, mainly in Production Planning/Material Management modules. For the

past a year and half, I've been mainly working on the field related to Personalized Access to Open Corpus Educational Resources. My primary interests are personalized web, social navigation, visualization, adaptive systems and recommender systems. I'm currently working with Dr. Brusilovsky and also a member of PAWS (Personalized Adaptive Web Systems) research group.

Keisuke Inoue

School of Information Studies

Syracuse University

kinoue@syr.edu

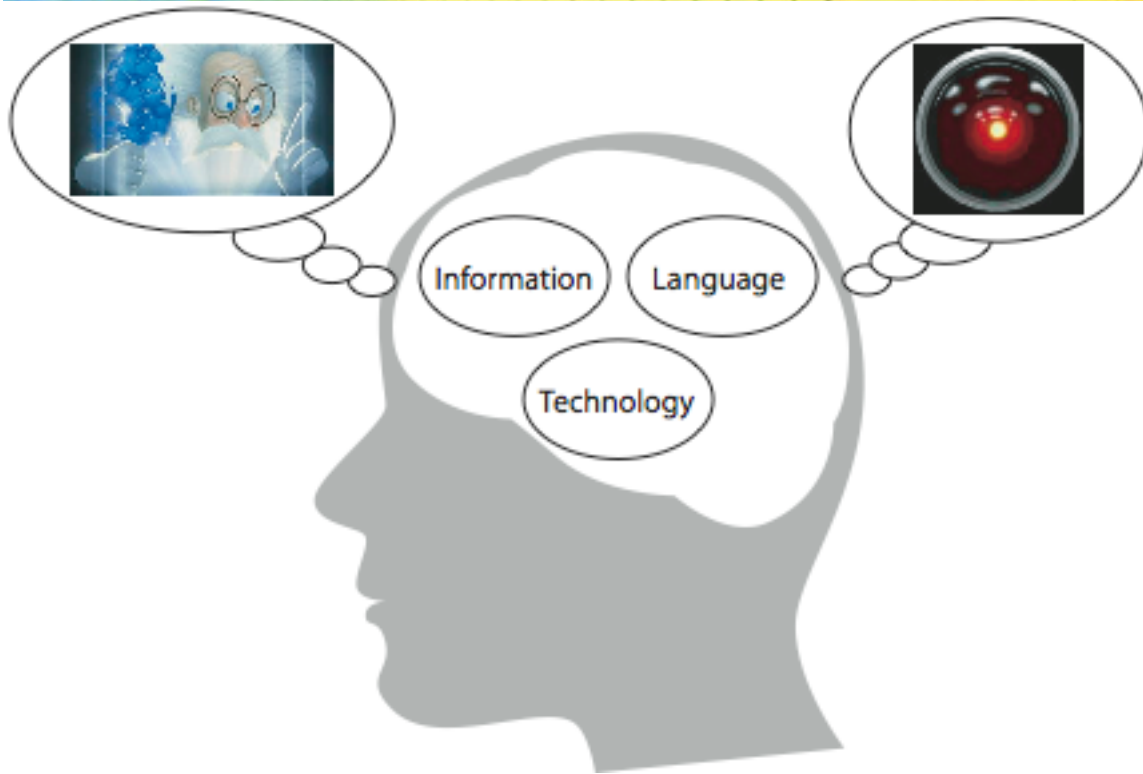


Poster title:

Conversation repository for participatory librarianship

Personal statement:

My primary areas of research are natural language processing, information retrieval and participatory network (and the intersection of all of the above). I am interested in pretty much anything that involves language, information and technology. I am also interested in systems with emerging behaviors: multi-agent systems, social networks, and of course, languages. My graduate education background is Linguistics and computer science. Like many other computer scientists, my lifetime research goals includes building something like HAL 9000 in the movie *2001: Space Odyssey* or Dr. Know in the movie *AI* (of course, not in a mad, human-attacking, way). That is, if I put it in the information science context, to design and implement an information retrieval system that facilitates interactions between the system and the user or among the users in order to enable various aspects of human information behavior: elicitation of the information need, feedback for the retrieved information, information use and exchange, knowledge creation and community building. My current research is to explore what is in human interactions, which enables those information behaviors. More specifically, I am looking at the linguistic nature of online conversations at various levels: syntax, semantics, and discourse, hoping I can find good characteristics and come up with implications for designing the information systems of the future!



Lian Jian

School of Information
University of Michigan
ljian@umich.edu



Poster title:

Why leave Wikipedia?

Personal statement:

I am interested in people's strategic interactions with information systems. I also love to help system builders design user contributed content systems. My academic background includes training in both electrical engineering and economics. I am looking forward to an academic career.

Xiaozhong Liu

School of Information Studies
Syracuse University
xliu12@syr.edu



Poster title:

Proposal of document classification with word sense disambiguation

Personal statement:

My name is Xiaozhong Liu, a PhD student from School of Information Studies, Syracuse University. My research areas are text mining, information retrieval and natural language processing. My undergraduate major is Computer Science, and I have three years working experience at telecommunication and software companies.

My research is focusing on three different but related areas. First is lexical level word similarity based on statistical and semantic approaches, which could be useful for query log analysis, word clustering and document classification. Instead of just focusing on traditional method, such like WordNet, I'm also trying to use the new lexical resource, such like Wikipedia, to identify the relatedness between words, phrases and proper nouns. Syntagmatic and paradigmatic relationship from corpus is also important in my research.

Second, I'm very interested in personalized IR systems. I do believe that the next generation IR systems will be based on personalized technology and personalized IR system will be increasingly important in people's life in the coming years. From user side, by analyzing and extracting users' knowledge (such as query log, selected documents or bookmark), we could provide much accurate IR results targeting on the specific customer. On the other hand, from the system view, we could build knowledge-based semantic indexing for document collection (based on domain ontology). The new system and algorithm could match the knowledge of users and documents rather than just keywords. Currently, I'm focusing on user blog and community based social network as the search context.

Last but not least, I'm interested in "Hot Term" prediction research, namely, use historical data, such as query log, selected document and daily news, to predict future user's interest. This method has granted a patent. "Hot term" prediction could be useful for IR ranking (today + tomorrow probability of relevance) or user based name entity identification.

My career goal is to be a faculty member in information retrieval and text mining. In the age of internet, we have got increasing amount of information, however, not enough knowledge. "Text mining" and "User study" could be our two important keys to answer this question. I will try my best to contribute to our research environment and teaching as well.

Alice E. Marwick

Department of Media, Culture, and Communication
New York University
alice.marwick@nyu.edu



Poster title:

To catch a predator? The MySpace moral panic

Personal statement:

Alice Marwick is a PhD Candidate at New York University in the Department of Media, Culture, and Communication. She studies social media from a critical perspective, including the political economy of new media, online identity, and feminist media theory. Her primary analytical lens is to view internet applications as commercial structures: how the business model of a website affects feature development, specifically the types of user actions that are encouraged or discouraged. Much of this work emphasizes how values often associated with social media, such as social interaction, privacy, participation, and creativity, must be understood within a commercial context. Her paper for the iConference addresses the “moral panic” around the social networking site MySpace.com and its purported link to “child predators,” linking proposed legislation to previous internet content legislation.

Recently, Ms. Marwick won the Association for Internet Researchers 2007 Student Paper Award for “The People’s Republic of YouTube? Interrogating Rhetorics of Internet Democracy.” This project investigated claims of YouTube as an inherently democratic space through an examination of community norms and Google’s corporate practices. Previous work on social networking sites such as MySpace, Friendster, and Orkut discussed how such sites typically circumscribe user actions, such as page customization, based on what is economically beneficial for the parent company. Ms. Marwick has also written about plastic surgery, reality television, identity management software, video games, online personals, profile customization, video-sharing sites, and desktop widgets. She has a MA in Computer-Mediated Communication from the University of Washington, where she was a fellow at the Center for Internet Studies. Her Master’s thesis, “Selling Your Self,” discussed identity presentation in commercial internet applications. Recent conference presentations include Association for Internet Researchers Conference (2005, 2007), the Microsoft Social Software Symposium (2007), South by Southwest Interactive (2007, 2008) and the Interfaculty Initiative in Information Studies Workshop at the University of Tokyo (2007).

Ms. Marwick is currently in the proposal stage of her dissertation: “Becoming Elite: Status in Social Media.” This research takes an anthropological approach to examining the creation and maintenance of status hierarchies within social media through ethnographic examination of workers in the Web 2.0 startups in the San Francisco Bay Area. Ms. Marwick is interested in how social media technologies contribute to status inequalities, particularly around gender, educational level, and class. The project is driven by four primary concerns: first, to examine how the features of certain interactive internet applications, both past and present, affect how status is maintained, displayed, or resisted; second, to determine the relationship between status hierarchies in “online” and “offline” social spaces used by the same community; third, to understand how status hierarchies affect the development and creation of social media applications; and fourth, to examine how status markers, particularly consumption, function in internet environments.

Alice has worked in the technology industry since 1995 as a researcher, writer, and strategist in fields such as video games, wireless, social networking, and software development. As a consultant, she is available for studies of user practice, online trends, and emerging technologies, particularly with regard to formulating recommendations for project design. Upon completion of the PhD in 2010, Ms. Marwick plans a career in

academia, with the additional goal of creating dialogue between technologists in private industry and technology scholars in the academy.

T. Patrick Milas

College of Information
Florida State University
pmilas@fsu.edu



Poster title:

From cultural participation to information visualization: A new framework for old knowledge management schemas

Personal statement:

Research Interests

My principal research interest is in cross-cultural, interdisciplinary theory development and application. The practical aim in my approach to theory development is to more closely map knowledge classification (or “KM”) schemas and information access protocols to user-defined relevance and information visualization applications. My three key research areas include: (1) Knowledge Management; (2) Information Needs, Seeking, and Use; and (3) Information Access Design and Policy.

The emphasis of my research is on how ancient knowledge management schemas inform the classification and information access practices of the discourse communities of world religions. My research analyzes depictions of identity formation in light of what I have termed the “faith-informed” face-to-face and virtual communities of cultural practice, particularly Orthodox Jews and Hindu tantrists.

I have used several theoretical frameworks in analyzing the socio-cultural associations between ancient texts and contemporary knowledge management. Elfreda Chatman uses the term “Small Worlds” to describe what she sees as the basic unit of social information behavior. My own research suggests that, while many quests for knowledge *qua* meaningful information serve a social function, others have a more dualistic purpose, displaying a tension between polarities (i.e. mercy and judgment, male and female, manifest and tacit).

My particular focus is on identifying and analyzing those works in which there is a striking balance between the dual elements of knowledge architecture – that is, systems which accommodate content equitably and organize it by polar contingencies (e.g., legal dispensations of criminal law [tractates] located opposite charity and social justice resources). Current research entails the use of content analysis on data (1) collected from survey mystical texts and intensive interviews of their practitioners, and (2) present in the blogs of the “Small Worlds” that are based on reference queries about Jewish law on AskMoses.com.

My research also addresses information policy and E-government evaluation. I make use of Brenda Dervin’s Sense-Making paradigm as a framework for understanding the power structure between government and citizens, and the libraries that increasingly serve as intermediaries for year-round provision of government services in electronic form (E-government), as well as emergency provision during disaster situations.

Background

During undergraduate research travel, funded by the Abernethy Research Fellowship for Comparative Religion, I conducted interviews of religious minorities in Austria, the Czech Republic, Germany, Hungary, India, Israel, the Netherlands, and Poland. To corroborate the qualitative data I gathered in my field research, I researched the archives of the Terezin Institute in Prague and the *Bibliotheca Rosenthaliana* in Amsterdam.

During my M.S. degree work in the College of Information, I worked in London, UK in information systems and technology. Later, as Doctoral Research Associate I led the Information Use Management and Policy Institute's IMLS grant writing (for \$249,000) regarding rural, marginalized communities e-government access during natural disasters (Co-PIs: Charles McClure and Jon Bertot). In order to address the national nature of the E-government delivery I used the element of demographics from Johnson's model in the sampling approach of libraries by geographic distribution, socio-economic indicators, and other variables such as average age, race, and distance from a public library – all characteristics that may exacerbate information poverty and indicate an elitist inattention to marginalized communities. After data collection, we will evaluate libraries', agencies', and patrons' insight into the information design of E-government. We aim to identify strategies transferable across emergency/natural disaster scenarios.

As a Doctoral Candidate I am currently collaborating in writing and managing the application process for a National Science Foundation (NSF) grant (for \$498,000) pertaining to the underrepresentation of Latinas in the IT workforce (Co-PIs: Kathy Burnett, Gary Burnett, and Paul Jaeger). Both projects propose to use intensive interviewing with minorities – a data collection method I employed when I interviewed holocaust survivors in Amsterdam, Prague, and Jerusalem as part of my George Abernethy Research project (for \$13,800) "Jewish Renewal After Nazi Occupation in Prague and Amsterdam: From Holocaust Survival to Religious Fervor" (PI: Patrick Milas).

A few of the most relevant ways my research connects to practice are: my presentation at the 2008 ALA Forum on Education and participation in sessions with practitioners; my conference paper presentation on "Outlet, Faceted Classification, and Theological Librarianship" to the practitioner community of the ATLA; my library use and scholarly communication with librarians in Amsterdam, Prague, and throughout Israel; my attendance at the Library Research Seminar IV in London, Ontario, where my IMLS-funded coding and analysis of virtual reference transcripts on the Florida Electronic Library was presented; and, my service for the conference committee for the information architects at the 2007 IA Summit of ASIS&T.

I have engaged issues about information in society by teaching about the social nature of reading in "Information Needs of Adults" and about library activism during World War I in the "History of American Libraries." My research on information in society has received several awards, including the Harold Lancour Scholarship for Foreign Travel.

Future Plans

After defending my dissertation in 2008, I will assume a position as a faculty member in an information school (iSchool). I chose to conduct doctoral research at an iSchool because of a longstanding career interest in a faculty position at an iSchool, though I am also interested in library research, hence my participation in both the faculty-oriented Doctoral Workshop at the Library Research Seminar IV (London, Ontario), and the iSchools' Doctoral Colloquium at the 2008 iConference at the University of California at Los Angeles.

Heather Lea Moulaison

School of Communication, Information and Library Studies

Rutgers University

heather.moulaison@rutgers.edu



Roundtable title:

The globalization of education for digital librarianship: Implications for iSchools in North America

Personal statement:

Heather Lea Moulaison is currently a doctoral student in Library and Information Science at Rutgers University. Before undertaking full time study, Heather worked in libraries as Cataloging/Modern Languages Librarian at The College of New Jersey (2005-2007) and as Cataloging Librarian at Missouri State University (2002-2005). Heather has been teaching in LIS programs (both at Rutgers and through a cooperative program with the University of Missouri-Columbia) since 2003. Before earning her MSLIS from the University of Illinois, Urbana-Champaign, Heather also earned a MA in French from the same institution. She has spent a total of four years studying and teaching in France, and speaks French fluently.

Heather's scholarly interests include the globalization of digital information, international librarianship, and Library 2.0. In December 2007, Heather published an article on Web 2.0 in France in *The French Review* and in October 2007 presented as part of a panel on communities in social tagging at the American Society for Information Science & Technology (ASIS&T) Annual Conference in Milwaukee. Her current research includes the study of social tagging and blogging, especially in France, author-generated indexing on the Web, next generation OPACs, user-centered design, and citation analyses of French negritude writers.

Heather has presented at numerous local, state, national, and international conferences. In 2007, she co-presented at the Association of College and Research Libraries (ACRL) 13th Annual Conference on the topic of library RSS feeds, gave a talk on Library 2.0 at the Ecole nationale supérieure des sciences de l'information et des bibliothèques (enssib) in Lyons, France, presented one of ten selected posters at the Association for Libraries and Technical Services (ALCTS) 50th Anniversary Conference in Washington, D.C. reporting on a study of search logs from a college library, and was able to participate in other local and regional conferences and events as well.

In 2008, Heather will be participating in the iSchool's iConference 2008 in February/March, and co-presenting at EDUCAUSE's NERCOMP (North East Regional Computing Program) 2008 Annual Conference in March, the Joint ASTED/CBPQ Conference in Montreal in May, and the 10th International Conference of the International Society for Knowledge Organization (ISKO) in August.

Jinfang Niu

School of Information

University of Michigan

niujf@umich.edu



Poster title:

The impact of documentation on secondary data use

Personal statement:

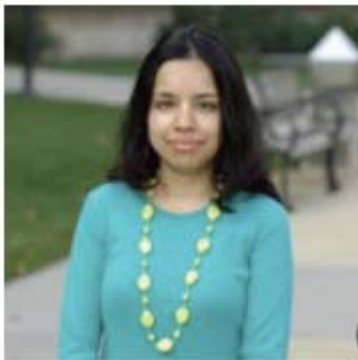
Jinfang Niu is a Ph. D candidate at the School of Information, University of Michigan. Her current research is about data sharing, secondary data use and data archives. She has work experiences and research background in metadata, digital libraries and digital preservation. Given her interdisciplinary interests, she has taken great effort in bring theories and methodologies in business management, social psychology and economics into the studies of archives and libraries.

Sharoda A. Paul

College of Information Sciences and Technology

Pennsylvania State University

spaul@ist.psu.edu



Poster title:

Healthcare informatics: Supporting collaborative sensemaking in the emergency department

Personal statement:

I am a 4th year Ph.D. candidate at the College of Information Sciences and Technology at Pennsylvania State University. In 2004, I received my undergraduate degree in Computer Sciences and Engineering from the Birla Institute of Technology, India.

My research interests broadly lie in the fields of computer-supported cooperative work, human-computer interaction, medical informatics, and information sciences. More specifically, I am interested in studying how groups use computers for collaborative information seeking and sensemaking.

My Dissertation Research: Supporting collaborative sensemaking

I am fascinated by the problem of how people make sense of large bodies of information. Often, this process of *sensemaking* requires pooling of information from various sources and media, as well as collaboration with

others. Specifically in the context of collaborative work, sensemaking is a complex process where the sensemaker's *understanding* of a situation is generated by interaction with others and maintaining awareness of the situation and of other's activities. *Collaborative sensemaking* plays an important role in everyday work as well as in crisis situations like emergency response. I am interested in examining how technology can be designed to support this process. Hence the research question I am interested in answering is:

“How can we support sensemaking in co-located groups engaged in information-intensive collaborative tasks?”

I have been conducting an ethnographic study of the sensemaking activities of healthcare providers at the emergency department of the Penn State Hershey Medical Center, a 500-bed teaching hospital associated with Penn State University. The emergency department is an ideal environment to study collaborative sensemaking. A challenging aspect of the work of doctors, nurses, and other staff is making sense of dynamic and unfamiliar situations in collaboration with team-members and using an array of information sources. In my data collection I have been specifically focusing on the *triggers* of collaborative sensemaking, the *challenges* to the process, and the *aids* used. The findings will inform the design of tools that can enhance collaborative sensemaking in groups.

Other interests include encouraging women to take up technology careers and exploring the role of technology in developing countries. In my free time I enjoy running, racquetball, skating, cooking, the theatre, and traveling.

Joseph B. Rubleske

School of Information Studies

Syracuse University

[jrubesk@syr.edu](mailto:jrublesk@syr.edu)



Poster title:

Towards a model of determinants of web services platform adoption by complementers

Personal statement:

The first substantial web application I developed was a fantasy basketball site in 1999, when the web itself still served as the favored application platform. (More on that issue below.) At the time I had just moved with my wife from Indianapolis to Madison, Wisconsin to live near family. In Indianapolis I had been a researcher at Indiana University's Center for Urban Policy and the Environment (CUPE) from 1995 through 1999, where I conducted economic development and local planning studies and assumed responsibility for designing and managing small project databases. From 1997 to 1998 my interests gradually shifted from economic development and planning to database management and web development. By late 1998 I decided I wanted to get paid to develop web sites, though I wasn't yet sure how to accomplish this.

In the summer of 1999 we relocated to Madison, and I gave myself a few months to learn how to build data-driven web sites. To demonstrate my ability to employers I wrote the back- and front-end code for a fully operational fantasy basketball site. (I'm an avid sports fan.) I presented the site in three job interviews and received two offers, accepting a web development job with the Wisconsin Department of Natural Resources.

In less than six months I was promoted to help develop an enterprise-wide environmental management system (EMS).

After training to use Oracle, I soon learned that the work involved less coding than expected. Instead, I spent much of my time in meetings with users to develop business requirements and translate them into conceptual and logical E-R models. In hindsight, this was an especially fruitful period of professional development for me, as I learned a great deal about project management and organizational behavior.

In late 2001, after more than two years of web development and database design, I returned to the research profession, accepting a more lucrative job as a senior research analyst at another state agency, where I evaluated the efficacy of program initiatives and developed requirements for internal data warehousing systems. In 2002 I concluded that an academic career was what I most wanted – just as my mentor, Dr. Samuel Nunn, had predicted back in 1998 when I left CUPE to pursue web development work.

The Doctoral Years (2003-present)

I sought admission to an information studies program – rather than an urban studies or computer science program – because the faculties at information schools were drawing from a range of discipline-specific knowledge – in areas such as management and organizational studies, information science, psychology, economics and information policy – to try to describe how information is created, destroyed, structured, managed and mismanaged and used and misused, and how these processes impact individuals, organizations and society. I was (and am) excited by the information field's open-endedness and relevance to contemporary problems.

I arrived in Syracuse in August 2003 expecting to apply quantitative methods to knowledge management research. One year later, I was convinced that I was going to write a dissertation on an ethnographic study of classification behaviors in large organizations. In my third year, my prospectus proposed a research design for investigating organizational change in an architectural firm that was in the process of adopting a disruptive technology. The firm's principals were supportive of the research until they decided that my presence – or even surveys I might administer – would promote unwanted tension.

While I learned a great deal from these and other forays, I realized in Spring 2007 that I had to commit to a topic. As alluded to in the first sentence of this statement, I am interested in the implications of the corporate 'platform-ization' of the web. The trend now is clearly toward development not on the web itself as a platform, but rather on the platforms of Google, Facebook, Amazon, eBay and others. There are no doubt cases where building on Facebook's platform (or storing your data on Amazon's servers, etc.) makes sense on some strategic and economic levels, but there are also potentially catastrophic repercussions from doing so. What happens when the platform owner is purchased, or changes its terms and conditions? What happens when your application becomes so successful that the platform owner sees fit to make a knock-off and squeeze you out of the market? With which companies is it sharing your data? What will happen when you need to migrate the data to another platform? With these and other questions in mind, my dissertation proposal involves interviews with owners of startups that build applications for these platforms or otherwise rely on them in some essential way. The research thus draws from concepts in risk analysis, contingency and strategic planning, entrepreneurial success and failure and web commerce.

In addition to my dissertation research, I have worked since 2005 on an NSF-sponsored study aimed at assessing the usefulness of genre information in web search systems. I am currently helping to design an experiment for this project and nearing completion of a chapter that will appear in a forthcoming book titled *Genres on the Web: Computational Models and Empirical Studies*.

Finally, I am teaching a graduate course titled Management Principles for Information Professionals for the fourth consecutive semester. It's a challenging course to teach, I believe, but a rewarding one. I strive to make the course as experiential as possible, supplemented with case studies and personal anecdotes from my work experience. My goal in the coming year is to complete my dissertation and interview (at AMCIS, ICIS and ALISE) for tenure-track positions at iSchools and business schools with IS departments.

Terrell Russell

School of Information and Library Science
University of North Carolina at Chapel Hill

unc@terrellrussell.com



Poster title:

Watching organizational opinion via social tagging

Personal statement:

Terrell Russell is a PhD student and teaching fellow at the School of Information and Library Science at UNC-Chapel Hill.

Terrell has B.S. degrees in both computer engineering and information technology and service organizations from North Carolina State University. He also has his M.S. in computer networking from NC State.

His research interests revolve around identity and reputation, social tagging and folksonomy, and cognitive authority and expertise. His planned dissertation topic is contextual authority tagging. He wants to investigate whether a group of people can sufficiently agree on what a group member knows. If this turns out to work in the physical world, he'd like to see if it works in a mediated space (online). If so, expertise location and the wisdom of the crowds could be combined to change the way we assess reputation and confer expertise.

He is a co-founder and developer of claimID.com. He runs the tagging site Cloudalicio.us. He is very interested in online identity and data portability as well. He would like for everyone to have the option of hosting their own data about themselves (mail, photos, health information, financial data) and share it on their own terms.

Terrell writes This Old Network at weblog.terrellrussell.com.

Frederic Stutzman

School of Information and Library Science
University of North Carolina at Chapel Hill

fstutzman@gmail.com



Poster title:

Longitudinal analysis of freshman adoption of Facebook.com

Personal statement:

Fred Stutzman is a Ph.D. student at the University of North Carolina at Chapel Hill's School of Information and Library Science, and co-founder of ClaimID.com, the identity-management site. His research interests include social computing, online social networks and digital identity. Prior to entering graduate school, Fred worked for Ibiblio.org (formerly SunSITE), the large digital repository of open-source, open-access content. While at Ibiblio, he proposed and managed the development of Lyceum, the open-source blogging platform. He has previously worked in technical and project management roles for The Motley Fool and Nortel Networks.

Fred's research on social networks and identity has been widely cited by local, national and international media. He speaks frequently about his research, presenting to organizations such as Google and Yahoo. He was the organizer of the UNC Social Software Symposium, and BarCampRDU, the tech unconference. He is currently co-organizing the upcoming ASIST Social Software Summit.

In addition to his academic work, Fred provides consulting to a number of organizations, including media and software companies, non-profits and political campaigns. Previous clients include the presidential campaigns of John Kerry and Wesley Clark. Fred also writes frequently; he maintains the blog Unit Structures (<http://chimprawk.blogspot.com>) and is a contributing author to techPresident (<http://techPresident.com>), a group blog examining how presidential candidates use web technology. He is the author of an ongoing series of articles examining Web 2.0 technology, published by O'Reilly.

Xiaoning Sun

College of Information Science and Technology

Drexel University

xiaoning.sun@ischool.drexel.edu



Poster title:

The impacts of gender and initial social activities on trust and performance in CMC

Personal statement:

I am a Ph.D. candidate at college of IST at Drexel. I received my M.S in information management at Syracuse University. I also have earned a M.E. (2001) in Industrial Engineering in the School of Management and a B.L. (1999) in the School of Law at Tianjin University, China.

In terms of the research work I've been doing, from the beginning of my doctoral studies I was attracted to Human-Computer Interaction (HCI) and specifically computer-mediated communication (CMC), with an emphasis on communication technologies, gender, and cross-cultural issues. Over the last three years I carried out a pilot study and two subsequent studies on gender effects in CMC technologies, culminating in my dissertation work. Gender differences are well known in the communications field, but I focused on aspects that have been less studied □ the effect of gender on carrying out complex, collaborative tasks in a virtual environment, mediated by modern communication media. The contribution of my work is to bring the gender factor into CMC from an HCI perspective, responding to central concerns of effectiveness, efficiency and user perceptions.

The primary objective of my dissertation work is to explore gender differences on trust and performance in synchronous computer-mediated technologies with and without initial social activities. The results from both quantitative and qualitative analyses suggest that gender as an important factor influences expectations and perceptions of communicators in CMC, in which female pairs had high levels of trust and more collaborative behaviors than male pairs in doing the task. In addition, females' collaborative conversational style focusing on harmonious relationships put them in a position to achieve trust in the communication. The results also suggest that initial social chat prior to beginning work help remote team members build trust in the communication, particularly, initial social chat is more effective in female dominated work groups. The results have implications for both the research and practice of building remote trust, as well as provide possible solutions to the design of interactive systems.

Beyond my dissertation work, I have taken steps to continue my work in CMC post-dissertation. I have joined with researchers from Drexel University, Long Island University, and Institute of Psychology, Chinese Academy of Sciences to carry out a cross-cultural study of CMC and gender, comparing pairs in the United States and China. This ambitious study adds a new dimension to my work and suggests future research directions.

In conclusion, I have developed a growing stream of research that I have built upon consistently over the last three years. In doing so, I have reached out to develop fruitful collaborations with colleagues even outside Drexel University. I have no doubts that I'll continue to make significant contributions to my research field.

I have a variety of interests besides the school work. I love to travel and watch movies; in particular, I specialize in music. I have played the violin for almost 30 years. This endeavor has taught me great discipline that I now apply it to all my assigned tasks. I've ever been a member of the Syracuse University Orchestra.

I am finishing my dissertation at Drexel, and now I am looking for tenure track faculty positions. I enjoy doing research and teaching, and would like to have more opportunities to collaborate with other researchers. I believe I will continue my success in conducting research with my persistent enthusiasm and efforts.

Rick Wash

School of Information
University of Michigan
rwash@umich.edu



Poster title:

Incentive design for home computer security

Personal statement:

Research Interests

My research interest is studying incentives in social software systems. It can be difficult to predict or understand user behavior in social systems because this behavior strongly depends on both the actions of others in the system and the specific design of the system. My work looks at how users use these technologies, why they use them in this way, and how different incentive structures in these systems can produce different

user behavior. I am particularly interested in software systems where users all have different information, and appropriate sharing of that information is key to the success of the system. Such systems need to be carefully constructed so that users are motivated to share their information, and to share it in a way that others can use.

I focus on applied, real-world problems, primarily problems arising around computer security and social computing. My research is grounded in social science theory from economics and psychology, providing a framework with which I can generalize my results and apply them to other interesting situations. My strong background in computer science allows me to understand technical constraints, and to design and implement solutions. I also frequently use mixed methods in my research; I believe that a variety of methods can reveal different aspects of a problem so that I can develop a more complete understanding and provide better solutions.

Recent Projects

Home Computer Security (Dissertation Research)

One of the most daunting information security problems is home computer security. Hackers are increasingly exploiting the millions of these machines to build “botnets,” consisting largely of home machines that hackers compromise in order to remotely control them. This mixed method project focuses on users and usage of security technologies that protect home computers. We use qualitative interviews to understand the mental models of the threats facing home computer users. We are designing a new security technology that helps users to better secure their own home computer. This design is strongly grounded in social science theory; we use social psychology theories and economic modeling to understand incentives in the system, individual responses to them, and inter-individual strategic awareness and behavior. Finally, we use a human-subjects lab experiment to validate these design principles and evaluate this new technology.

Rick Wash and Jeffrey K. MacKie-Mason. “Security When People Matter: Structuring Incentives for User Behavior.” *International Conference on Electronic Commerce*, 2007.

Social Tagging

User-contributed metadata, also known as *tagging*, is increasingly receiving attention as a loweffort tool for digital information management. Collaborative tagging systems provide a means for users to associate personally salient keywords or labels with content items, and then publicly expose these associations, so everyone can benefit from this information. Most research in this area has focused on the outputs of these systems, such as the ‘folksonomy’ that emerges from many users making many tag choices. Instead, we focus on the inputs; how do users choose which words to use as tags and when to apply them to a content item? We also look at how systematic differences in this input -- tag choices -- lead to specific patterns in the output folksonomy, which after all is an aggregate of these tag choices. This mixed method project uses qualitative interviews, large-scale statistical analysis, and computer simulations to understand user behavior on a prototypical tagging system -- del.icio.us.

Rick Wash and Emilee Rader. “Public Bookmarks and Private Benefits: An Analysis of Incentives in Social Computing.” *ASIS&T Annual Meeting*, 2007

Rick Wash and Emilee Rader. “Understanding del.icio.us Tag Choice Using Simulations.” *iConference 2008*

Anti-spam

Spam email is fundamentally an economic problem. We characterize a number of existing technological and regulatory solutions to the spam problem in economic terms, and explain a number of their faults. We then develop and analyze a new anti-spam solution (the Attention Bond Mechanism) that uses economic incentives to discourage spammers. We take advantage not of the information contained in a message, but of the information known by the sender to screen email before it is even sent. This idea is currently being developed into a product by boxbe.com

Thede Loder, Marshall van Alstyne, and Rick Wash. (2006). "An Economic Response to Unsolicited Communication." *Advances in Economic Analysis and Policy* 6(1).

Thede Loder, Marshall Van Alstyne, and Rick Wash. “An Economic Solution to the Spam Problem,” *Proceedings of the ACM Conference on Electronic Commerce*, 2004

Education

- B.S. Computer Science 2002. *Case Western Reserve University*
- M.S. Computer Science 2005. *University of Michigan, Ann Arbor*
- Ph. D. Candidate in Information, May 2006. *University of Michigan, Ann Arbor.*
- Expected graduation: May 2009
- Advisor: Jeffrey K. MacKie-Mason

Qin Wei

Graduate School of Library and Information Science

University of Illinois at Urbana-Champaign

qinwei2@gmail.com

Poster title:

Interactive machine learning (IML) markup of OCR generated text by exploiting domain knowledge: A diversity case study

Personal statement:

Hello, Everyone, My name is Qin Wei, a Ph.D student at Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign (UIUC). My research interests are mainly information retrieval and information extraction, especially in automatic text information extraction by using Machine Learning (ML) and Natural Language Processing (NLP) techniques. I have a bachelor degree from Department of Information Management, Peking University and an Economics bachelor degree from China Center of Economics Research. After that, I came to UIUC for the Master degree and the Ph.D degree.

This is my second year at the Phd program and I am finishing my course work. I plan to take the qualify exam this coming fall semester and begin to prepare my dissertation proposal at the same time. My tentative dissertation topic will be noisy text classification by using interactive machine learning. The topic is close related to my current research work. I am currently working on Herbis (www.herbis.org) project, specifically, the NLP module of the whole project. Herbis (Herbis is the Erudite Recorded Botanical Information Synthesizer) is a system which offers 'one-button' specimen imaging and data capture. The system contains three modules: data capture (mainly image scan), Optical Character Recognition (OCR) and NLP parser. Till now, I have implemented two Machine Learning algorithms: Naïve Bayes and Hidden Markov Model as the parser. And I put a Herbis NLP Demo online which could be found here (<http://www3.isrl.uiuc.edu/~TeleNature/Herbis/src/web/htdocs/>). In the meantime, I am doing research on how to improve the performance of ML in the domain of automatic noisy text classification under the guidance of my advisor, Dr. Bryan Heidorn. Noisy text here means the output of OCR of not the perfect human type-in texts. We have given out several presentations and posters about our preliminary results in biological database conference and also digital library conference.

My current career goal is to be a faculty member in information schools in the US or China and continue my research on natural language processing. For this goal, the first objective I have now is to complete my dissertation and gain my Ph.D. degree at UIUC. During my time at UIUC, I aim to utilize my time as a

doctoral candidate to gain the skills and experience necessary to become a good researcher and professor, especially when my area is a highly interdisciplinary field. I need to keep myself updated frequently.

In closing, I look forward to meet all of you from different background and interests of information science. And I appreciate the opportunities provided by iConference'08 of presenting my own research and discuss it with experts in this field. Thanks for the supports from my advisor, Dr. Heidorn and other faculty members of UIUC. I also look forward to bringing the ideas get from this conference back to my research and studying.

Pengyi Zhang

College of Information Studies

University of Maryland

pengyi@umd.edu



Poster title:

Support sense-making with tools for structuring a conceptual space

Personal statement:

I am Pengyi Zhang, a fourth-year Ph.D. student in the College of Information Studies at the University of Maryland, College Park.

My research interests lie in the broad areas of information retrieval, information seeking and use, natural language processing, and cognition. Some sub-areas that I am interested in and have been working in include:

- Sense-making
- Interactive information retrieval
- Information seeking and use behavior and cognition
- User-centered design and evaluation
- Natural language processing
- Decision making

I hold a B.S. in Information Management and Systems (2003, Peking University) and a M.L.S. (2007, University of Maryland).

At present, I am working on my dissertation research. My dissertation is to design and evaluate sense-making tools that enable users of information systems to structure their problem spaces of a task through automatically extracted concepts and relationships and graphical manipulation of the conceptual space. The research aims to build on sense-making, cognitive psychology, and learning theories to investigate users' sense-making processes using the tools and to provide design implications.

To me, the ultimate goal of information systems is to help users find the information that assists them in successfully completing their tasks and solving their problems. Helping users retrieve the right information is only half the battle; assisting users with making sense of what they found is the next frontier in information system design. I hope that my dissertation research will contribute to our understanding of sense-making processes and tools and thereby give a better foundation for system design.

My career goal is to be an excellent researcher and a good teacher, to bring technology to users and to help my students learn.

Nan Zhou

College of Information Science and Technology

Drexel University

nan.zhou@ischool.drexel.edu



Poster title:

The collaborative information behavior of middle school students in online learning environments: An exploratory study

Personal statement:

I am currently a doctoral candidate at the iSchool of Drexel University. Prior to Drexel, I got my master's degree in Information Management & Systems from University of California at Berkeley. My concern is understanding the interaction between people and information in a networked social world connected by information and communication technologies. My research can be characterized as lying at the intersection of information science and Computer-Supported Collaborative Learning (CSCL). I am investigating how small groups in online learning environments make use of information resources to collaboratively build knowledge. My current research focuses on understanding information practices of online collaborative small groups, which helps us design information resources and environments to support group learning and collaboration.

In the past 4 years, I have been involved in a research project called the Virtual Math Teams (VMT) where researchers from different disciplines explore the design of learning environments to support people collaborating on doing math and to build an online math discourse community. I have been working closely with Dr. Gerry Stahl, my dissertation supervising advisor, the director of the VMT project. My research is mainly qualitative and draws heavily on ethnomethodology. I apply conversation analysis techniques to carry out case studies and analyze episodes of interactional data in detail. I have taken an analytical perspective to study the phenomena of interest interactionally, treating the group as the unit of analysis. Some of the research topics include examining how people ask information seeking questions, how information needs are negotiated, and how information becomes meaningful in and through interactions of participants in the group. I have been exploring how such understandings can help us with virtual community building, designing digital libraries, and designing of information resources. Some of the research findings have been published and presented at some major conferences in the field including the Annual Meeting of ASIST, the Joint Conference of Digital Libraries (JCDL), the Interactional Conference of Computer-Supported Collaborative Learning (CSCL), the International Conference of the Learning Sciences (ICLS), etc.

Currently I am focusing on the data analysis for my dissertation work and at the same time writing some findings up for an edited book of the VMT project. I hope to finish my doctoral study by summer 2008. I am looking for a faculty position or other research position where I can continue my current research and take it further.

Full Papers

(33 Accepted Submissions)

FULL PAPER SUBMISSION 1:

Shaping Transformation or Continuity in the (Social) Sciences: The Internet and e-Research

AUTHOR(S):

Meyer, Eric T
Schroeder, Ralph
Dutton, William H.

ORGANIZATION(S):

Oxford Internet Institute, University of Oxford, United Kingdom

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

There are major efforts underway in the United States, United Kingdom, Europe, and elsewhere to build scientific infrastructure to enable e-Research. In the U.S., the NSF Office of Cyberinfrastructure is engaged in an ambitious programme aimed at encouraging and developing cyberinfrastructure in a variety of science and social science domains (see, for example, <http://www.nsf.gov/od/oci/reports.jsp>). In the U.K., the ESRC National Centre for e-Social Science (NCeSS) is currently awarding its second round of funding to projects aimed at understanding the requirements for grid-enabled e-Social Science and developing tools for e-Research (National Centre for e-Social Science, 2007). Other specific grid-based computing projects aimed at enabling e-Science and e-Social science projects abound in the U.S., U.K., E.U., and elsewhere.

Much of the language about these new scientific information infrastructures stresses that many funding bodies view e-Research as a way to transform scientific practice. The National Science Foundation (NSF), for instance, has recently revised “the intellectual merit review criterion to specifically include language on potentially transformative research” (National Science Foundation, 2007). Although language indicating the potential for e-Research to be transformative abounds, very little work has examined the extent to which this potential is being realized.

In addition, the notion of transformation in e-Research can operate on many levels: transformations in the tools used to conduct research, transformations in projects enabling e-Research as they develop, transformations to ordinary scientific practice, transformations in the types of scientific questions asked and able to be asked, and transformations in the scientific imagination. While much of the rhetoric implies that e-Research will transform the very nature of science, other types of less-pervasive transformations are more evident at these relatively early stages in the development of e-Research infrastructures.

To examine this concept of transformation, the authors of this paper are currently in the process of conducting case studies in four U.K. e-Social Science domains. This paper will report on the progress of this study and present initial findings and results from these cases. Each of the case studies involves examining one or more specific projects engaged in developing tools to enable e-Research in a specific domain. Each of these specific projects constitutes a sub-case of the main domain-specific case. The four domains/case studies being examined are qualitative social science, quantitative social science, geospatial modelling and simulation, and web 2.0 collaborative e-Science. The methods for this research are a mix of qualitative and quantitative methods, including interviews both with project personnel and the users of their scientific tools, observation of how the tools are implemented, analysis of supporting documentation, and a quantitative survey of U.K. social scientists about their use of e-Social Science tools and their attitudes toward e-Social Science in general.

One of the challenges to understanding the nature of how e-Research is developing is that many of the projects developing e-Research tools have, until fairly recently, been primarily developing tools that hadn't yet been released or tested widely in real-world settings. As the efforts have matured, however, it has become feasible to study these projects and to begin developing an understanding of the social shaping that has occurred, and the social, organizational, ethical and legal issues that are associated with the projects. For each of the domain-level case studies to be reported in this paper, the sub-cases are all projects that have advanced far enough beyond their early stages to have created tools that are being used by scientists. In this way, we are able to examine transformations in tools, projects, scientific practice, and the formulation of scientific questions. In some areas, early results indicate that many transformations have taken place. For instance, at the micro-level, many of the projects have experienced significant transformation from their earliest roots as the initial ideas were shaped by experience. As Schroeder & Fry have observed, “the obstacles to the effectiveness of e-science are not so much technical, as social” (2007, p. 1). These social barriers are part of the feedback loops that operate as the projects transform over time.

At other levels, however, there is currently less evidence that e-Research tools and e-Infrastructures have yet transformed either meso-level practices of ordinary science or macro-level questions that science is able to imagine and study. One question we will grapple with in this paper based on the evidence we have collected is whether this lack of fundamental transformation is a temporary problem which will disappear as tools become better developed and more widely used, or whether there are larger issues that make it unlikely that science will be transformed so much as reinforced by e-Research tools. If the former is true, then it becomes just a matter of time and effort to reach the critical mass required for transformation. If the latter is true, however, then one needs to examine the underlying forces that serve to resist change in the fundamental science and instead co-opt new tools to the service of large, well-funded projects at the expense of smaller and possibly riskier areas of science. Indeed, it is very difficult to predict in advance whether new scientific tools are likelier to support normal science or to enable paradigm shifts, to use the language of Kuhn (1962).

We also are examining the tensions that have arisen surrounding the cases and sub-cases in the study. The particular tensions of interest are the social, legal, ethical, and organizational tensions that come about as the projects and domains have worked on developing effective infrastructures for e-Research. Social tensions have developed as competing parties with differing visions of the future of their scientific domains have negotiated the features and capabilities of e-Research infrastructures. Legal issues such as intellectual property concerns underlie many of the projects and many legal disputes have arisen that have had to be or are still being negotiated. Organizational and domain issues include the assignment of scholarly credit and the resulting rewards for contribution to shared data structures, and issues of control within domains and organizations.

This paper is designed to contribute to a dialogue about the future of collaborative scientific infrastructures in the U.S., the U.K., and elsewhere. By presenting these British cases to the mostly American audience represented at the iConference, we hope to be able to discuss similarities and differences between the U.S. and U.K. experiences and to use these discussions to develop a better understanding of the general issues underlying e-Research.

References cited

- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- National Centre for e-Social Science. (2007). NCeSS webpage. Retrieved October 26, 2007, from <http://www.ncess.ac.uk/>
- National Science Foundation. (2007). Proposal and Award Policies and Procedures Guide NSF 08-1, January 2008 (Document number nsf081). Retrieved October 29, 2007, from http://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp
- Schroeder, R., & Fry, J. (2007). Social Science Approaches to e-Science: Framing an Agenda. *Journal of Computer-Mediated Communication*, 12(2), Article 11.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

The Role of e-Infrastructures in the Transformation of Research Practices and Outcomes

Eric T. Meyer

Ralph Schroeder

William H. Dutton

Oxford Internet Institute, University of Oxford

1 St Giles, Oxford, OX1 3JS, UK

Tel +44 (0) 1865 287210

eric.meyer@oii.ox.ac.uk ralph.schroeder@oii.ox.ac.uk william.dutton@oii.ox.ac.uk

ABSTRACT

In this paper, we examine transformations that have taken place in e-Research, and address the potential for additional transformations as e-Research develops and matures. The notion of a transformation in e-Research can operate on many levels: transformations in the tools used to conduct research, transformations in projects that enable new types of e-Research, transformations to ordinary scientific practice, transformations in the types of scientific questions that can be asked and able to be asked, and transformations in the scientific imagination. While much of the current rhetoric implies that e-Research will transform the very nature of science, other types of less-pervasive transformations are more evident at these relatively early stages in the development of e-Research infrastructures, and some evidence supports the idea that continuity has been more common in the ordinary scientific practice of e-Research rather than transformation. The data from this paper draws on the work of the Oxford e-Social Science project (OeSS).

Topics

Information infrastructure development
Information policy, ethics and law

Keywords

e-Research, e-Social Science, scientific transformation, information infrastructure

1. INTRODUCTION

There are major efforts underway in the United States, United Kingdom, Europe, and elsewhere to build scientific infrastructure to enable e-Research. In the U.S., the NSF Office of Cyberinfrastructure is engaged in an ambitious programme aimed at encouraging and developing cyberinfrastructure in a variety of science and social science domains (see, for example,

<http://www.nsf.gov/od/oci/reports.jsp>). In the U.K., the ESRC National Centre for e-Social Science (NCeSS) has recently awarded its second round of funding to projects aimed at understanding the requirements for grid-enabled e-Social Science and developing tools for e-Research [19]. Other specific grid-based computing projects aimed at enabling e-Science and e-Social science projects abound in the U.S., U.K., E.U., and elsewhere.

Much of the language about these new scientific information infrastructures stresses that many funding bodies view e-Research as a way to transform scientific practice. The National Science Foundation (NSF), for instance, has recently revised "the intellectual merit review criterion to specifically include language on potentially transformative research" [20]. Although language indicating the *potential* for e-Research to be transformative abounds, very little work has examined the extent to which this potential is being realized.

The notion of a transformation in e-Research can operate on many levels: transformations in the tools used to conduct research, transformations in projects that enable new types of e-Research, transformations to ordinary scientific practice, transformations in the types of scientific questions that can be asked and able to be asked, and transformations in the scientific imagination. While much of the current rhetoric implies that e-Research will transform the very nature of science, other types of less-pervasive transformations are more evident at these relatively early stages in the development of e-Research infrastructures. In this paper, we examine some of the transformations that have taken place in e-Research, but we also address the potential for various transformations as e-Research develops and matures.

2. SOCIAL SHAPING OF E-RESEARCH

2.1. OeSS: Oxford e-Social Science Project

The Oxford e-Social Science Project (OeSS) is a node of the National Centre for e-Social Science (NCeSS). NCeSS was established with UK Economic and Social Research Council (ESRC) funding in April 2004 "to act as the central resource base for e-social science issues and activities in the UK...[and to] provide a one-stop shop for awareness raising, expertise, training, technical infrastructure, data resources, computer facilities and user-support for e-social science research" [3]. To do this, a central hub was established at the University of Manchester, and eight individual nodes were funded in 2005 for an initial three year period; NCeSS also funded twelve smaller pilot projects that ran for short periods ranging from 9-18 months. In 2008, a second round of three-year funding was announced which included continuing support for the OeSS node through 2011.

Seven of the initial nodes are primarily concerned with creating demonstrator projects which show the potential of e-Social Science in a variety of domains. The OeSS node is somewhat different than the others: rather than creating software tools to enable e-Social Science, the OeSS node studies the social shaping of e-Research. In particular, it focuses on the social, institutional, legal, and ethical contexts that shape the development and adoption of new technologies. It does this work by examining how various e-science and e-social science projects develop and change over time, including cases drawn from the other NCeSS nodes and cases from outside of the NCeSS project.

2.2. OeSS Cases and Themes

The OeSS project is currently in the process of conducting ongoing case studies in several e-Research domains. Each of the case studies involves examining one or more specific projects engaged in developing tools to enable e-Research in a specific domain, and each of these cases also contributes to the investigation of a number of themes in specific areas (including ethics and trust, intellectual property, openness of access, and disciplinarity). An attempt is also made to distribute the cases across a wide range, in social science, this includes qualitative social science, quantitative social science, geospatial modelling and simulation, and web 2.0 collaborative e-Social Science. The methods for this research are a mix of qualitative and quantitative methods, including interviews both with project personnel and the users of their scientific tools, observation of how the tools are implemented, analysis of supporting documentation, and a quantitative survey of U.K. social scientists about their use of e-Social Science tools and their attitudes toward e-Social Science in general.

While these case studies are still ongoing, in this paper we will describe an illustrative case and discuss examples of some of the cross-cutting issues and themes that are emerging across a variety of cases. Finally, we will draw some initial conclusions about how this research helps to illustrate social shaping in e-Research.

2.2.1. Geo-spatial modelling and simulation: GeoVue

GeoVue is an NCeSS node funded for both phase I and phase II of NCeSS. The project is based at the University College of London's Centre for Advanced Spatial Analysis (UCL-CASA), where Mike Batty is the principal investigator. GeoVue's central focus is making tools that allow the fast and easy creation of maps that include data elements for visualizing geo-spatial data. One example of the node's output is the Virtual London demonstrator (see Figure 1), which offered users a virtual fly-through of London using the Google Earth viewer as a platform.



Figure 1: Virtual London (Source: Hudson-Smith [13])

When GeoVue created Virtual London, the target audience they had in mind was primarily urban planners and policy makers. Virtual London was designed not only to create a useful 3-dimension map of London, but also to allow overlaying data. For example, local governments could overlay pollution readings and visualize areas where pollution abatement measures are most needed. Once the application was completed the project attracted the attention of Google, who expressed an interest in including it in the Google Earth viewer so that users anywhere could have access to Virtual London.

It is at this point in the story that institutional and legal arrangements intervened and prevented the public release of Virtual London. The Virtual London application, while it was built in the Google Earth viewer, relied on the Ordnance Survey's *MasterMap* data to build the 3-D projections of buildings. Ordnance Survey is the national agency of the British government that is "responsible for the official, definitive surveying and topographic mapping of Great Britain" [21]. In order to include Virtual London in Google Earth, Google would have had to obtain permission to use the Ordnance Survey data underlying the application. Ordnance Survey data, however, is protected by Crown Copyright, which covers all works "made by an officer or servant of the Crown in the course of his duties" [2]. This is in sharp contrast to the situation in the United States, where "copyright protection ... is not available for any work of the United States Government" [1]. Because works created by federal agencies are not subject to copyright protection in the U.S., data created can be used by individuals and by organizations wishing to put it to use. In the U.K., that is not the case – the Ordnance Survey data must be licensed for use.

In the case of Virtual London, Google was willing to pay for a license, and indicated that the actual amount wasn't really too much of a concern. According to a GeoVue staff member, "Google were willing to pay whatever it took, yeah. That was the whole point. They would pay whatever Ordnance Survey said ... [But] Ordnance Survey wouldn't budge for any price" [Interview 3 Oct 2007]. The only possible price the Ordnance Survey was able to offer was £1.50 "per click" which was not possible in Google Earth where this type of clicking is not part of the interface. In Google Earth, the user zooms and flies (virtually) through the space. Neither Ordnance Survey nor Google were able to suggest an equitable way of measuring the uses of Virtual London. Even though later discussions took place among GeoVue, Google and the Ordnance Survey, in January 2008 a member of the GeoVue project team indicated that he had given up on holding out much hope for getting Virtual London released and had moved on to other projects [Personal conversation 16 Jan 2008]. It can be added that throughout the duration of the project, there was a high profile campaign led by The Guardian newspaper to open access to the Ordnance Survey and other government data, a campaign which is still ongoing¹. In other words, the issue of Ordnance Survey access was not limited to the GeoVue project, but part of a wider debate in the UK.

This case is a clear illustration of the legal and institutional issues that can erect barriers to sharing data and implementing e-Research projects. It is often assumed that many projects fail to

¹ See <http://www.guardian.co.uk/technology/freeworld>

achieve their goals because of technical barriers to implementation. In the case of GeoVue, however, the GeoVue team had successfully removed the technical barriers and had built a functioning, effective application with the potential to have an impact on public policy and academic research. Legal and institutional arrangements, however, prevented Virtual London from being released widely. The legal issues are based on national copyright requirements, as discussed above.

Institutional arrangements also came into play however. When GeoVue and Google approached the Ordnance Survey for permission to use the Ordnance Survey data that was used to generate Virtual London, they ran into considerable institutional indifference above and beyond the legal barriers. In other words, GeoVue felt that not only was there an issue surrounding copyright and payment for licensing, but also a feeling that Ordnance Survey just didn't think that such an application was of any interest to them. This became even clearer to GeoVue project members after they set up a sample Virtual London in their space on Second Nature Island in the Second Life virtual world. GeoVue felt that the Second Life Virtual London demonstrator would be seen as a feather in the cap of the Ordnance Survey, which was clearly identified as a source of data. The data was not exposed in the demonstrator so it was not making it possible for non-licensed users to obtain protected data, and GeoVue felt Ordnance Survey could use Second Life Virtual London to show people how their data could be used. They were taken aback, then, when the Ordnance Survey response was a demand to immediately remove the demonstrator or risk legal action. They had no choice but to remove it, and replaced Virtual London with another demonstrator, Virtual Phuket.

In this case, several institutional and disciplinary contexts came into conflict. GeoVue is comprised of academic geographers and computer programmers. Like most other academics, the notion of sharing one's results and publishing one's output without expectation of direct financial reward is part of their institutional and cultural expectations. They may also have been somewhat naïve about the legal issues involved, as the team did not include lawyers or copyright experts. There would have been little reason to expect to need to include such experts, when the initial plan was to figure out ways to combine computer programming power with questions of interest to geographers, government planners, and other academics. This lack of expertise, however, resulted in being caught unaware when they came up against the government bureaucrats and lawyers working for Ordnance Survey. Google's institutional culture as an American corporation also had an effect: Google's response was fairly typical of the stereotype of Americans, to simply assume that throwing enough money at the problem would cause it to go away.

While this case illustrates a constraint imposed by legal and institutional considerations, it nevertheless also shows some of the promise of transformation that projects such as GeoVue offer to social scientists. In the past, the use of geospatial information has been relatively limited because of the technical complexity of working with map data and GIS programs. Recently, however, and especially with the advent of the Web, there has been a surge of interest in digital mapping techniques and combining these with different types of data and information. For many social scientists who wanted to include geospatial analysis and data presentation, unless they had a GIS expert on staff or available to

them through their organization they would have faced too great a hurdle and often decided to do without the visualization component in their research. GeoVue and another NCeSS node, MoSeS both have been working to make the creation of map data much more transparent to the average scientist; in doing so, they hope to transform the way social scientists see their data. This potential transformation may be quite powerful, and the OeSS project will continue to follow developments in this area².

2.3. Cross-cutting themes and issues

The case studies in the OeSS project are designed not only to illustrate particular cases of interest, but also to illuminate our understanding of a variety of themes and issues that have emerged during the first phase of the research. These themes and issues are threads that run throughout the work, and help describe how e-Research is undergoing social shaping at this relatively early point in the development of many e-Research projects. Among the important themes are trust, privacy, disciplinarity, and openness. A number of other issues are not properly called themes, per se, but also represent important issues that occur in a variety of projects. These issues include the difficulty of sharing data, the limitations of geography, economic constraints, shifts in funding models, and ways of approaching new scientific questions. In the interests of limited space, we will touch on several of these only briefly here.

2.3.1. Challenges for e-Research

Data sharing

Data sharing has proven to be a significant challenge for many efforts at building collaborative infrastructures for research. Beyond legal and institutional barriers to sharing data, such as the GeoVue case described above, there are a variety of other issues that arise as researchers contemplate sharing data. OeSS project members Carusi and Jirotko, for instance, have published work discussing how a seemingly simple data archive can become an "ethical labyrinth" [7]. Digital archives, they argue, can help normalize scientific research by increasing accessibility, uniformity and transparency. Archived data can also help scientists avoid needless repetition, and enable secondary uses of data. However, in practice much social science data resists archiving, particularly qualitative research data. Issues such as guarding subject privacy through anonymization can render otherwise rich data essentially meaningless as important relationships are stripped from sanitized data. Other contextual information may not be apparent in the data, and secondary researchers may mischaracterize data in ways that the original researchers who were embedded in the social relationships under study would not do. Re-using data for purposes beyond those initially stated also has implications not only with regard to informed consent agreements, but also run the risk of jeopardizing the relationship between the original researcher and their subjects if the subjects feel misled or duped by the subsequent uses.

² The GeoVue and MoSeS nodes of NCeSS have merged their efforts for phase 2 funding. This will also necessitate some transformations in work practices, as the two nodes are located in different cities and have been targeting somewhat different audiences to this point.

Another major issue standing in the way of widespread data sharing is the unresolved issue of how academic researchers can be assigned credit and rewards for contributing scientific data to a public archive. The reward structure of academia currently favours publication in top peer-reviewed journals, and subsequently having one's work cited by other peer-reviewed articles. Most measures of academic quality offer no credit for having created a dataset that other researchers use, or for creating a software package that enables research. Indexing services such as the Web of Knowledge do not index databases. The promise of e-Research in this area represents a potentially major transformation in scholarly practice: should the publication of scientific data become a standard part of the scientometric measures that measure scientific impact and are often used in the evaluation of researchers, departments, and organizations (see also [5])?

Some efforts are being made to address this issue of academic credit for data sharing. OeSS researcher Meyer has studied an American biomedical project dealing with issues of sharing data, the Genetic Association Information Network (GAIN) [17]. GAIN is a public-private partnership that offered academic researchers high-quality genotyping an order of magnitude more detailed than they had previously used in exchange for releasing their data and blood samples for other researchers and private organizations such as pharmaceutical companies to use. The data distribution for GAIN is being implemented in a system called dbGaP by the National Center for Biotechnology Information (NCBI), a national resource established by the U.S. National Institutes of Health. Among the features of their data distribution site are ways of seeing the number of publications and the authorized data requests that have been made for data (see Figure 2).



Figure 2: Selections from NCBI dbGaP website
(Source: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>)

In Figure 2, we can see parts of the page for the Framingham heart study portion of dbGaP. By identifying the authorized data access requests and the publications associated with the study, the creators of the data begin to have a way to demonstrate the impact of their work creating the dataset. While the well-known Framingham study certainly has many other indicators of success available to them as well, other less well-known studies can track

who is using their data, often for the first time. While some of these datasets may have been shared and distributed on an *ad hoc* basis in the past, without a central resource to track data uses, knowing how one's data is used can be difficult or impossible. Furthermore, until resources are made available for professional management of such resources, most datasets will continue to remain scattered across individual researcher's hard drives and local networks.

Again, we see a potential for transformation in scientific practice here, one that is closer to realization through the combination of technological enablers (the dbGaP website), scholarly incentives (access to additional data), and economic incentives (access to additional future funding for scientific work).

e-Science as Open Science?

Another theme of the OeSS project has been the openness -- or otherwise -- of e-Research. One of the promises of e-Research has been that it should enhance open access to knowledge, and thus to "reconfigure access to knowledge" [10] in terms of allowing data and resources from e-Research to be shared and disseminated more freely, with fewer restrictions (for example, by means of copyrights and patents) and perhaps using open source licensing or similar. There has been widespread support from funding bodies, researchers, science policy-making bodies and non-governmental organizations for 'open science' and 'open access' policies and practices, but there has also been evidence of resistance from stakeholders such as publishers who do not want published material made available freely online and researchers who do not want deposit their data immediately to share it with others. On closer inspection then, the openness of e-Research, widely espoused as an aim, needs to be disaggregated and its potential and implementation in practice evaluated in a way that goes beyond the general principle.

To do this, we can take, first, the most macro-level of e-Research, the 'infrastructures' that e-Research is in the process of developing. On this level, an attempt is being made to create shared sets of tools and resources for researchers in and across certain disciplines that are akin to the social infrastructures (transport, communication) that have been created to support social development. This trend is particularly clear in the US, where the term cyberinfrastructure is widely used to describe developments in e-Research. Here the support of science policy makers and funding bodies has been critical, and much of the effort of national and international e-Research programmes has been devoted to infrastructure development. The efforts to make these infrastructures 'open', however, must be put into an even larger context of current conflicts over intellectual property regimes, and the forces that are aligned to change them towards greater openness - or to maintain the existing regimes of restrictions and access rights. These include commercial interests, NGOs, and policy makers, and it remains to be seen how the current interplay of forces will shift the momentum of infrastructural systems in one or other direction [22].

A second level are the policies and practices of individual e-Research projects: whether the materials relating to projects' ongoing research are deposited in a common or openly shared repository; how software tools are licensed (various options of open source licensing); and whether data or publications resulting from the project are made available on project webpages or other

shared repositories. A major push, especially from research funding bodies, has been to encourage 'open' practices. For example, an increasing number of funding bodies stipulate that funded projects must make their results freely available online (perhaps after an embargo period with exclusive rights to publish reserved for the contributing researcher). What we found in practice, however, according to in-depth interviews with 12 researchers on UK e-Science projects is that practices vary considerably across projects [11]. Many projects are uncertain about what the policies about copyright and licensing are; they often adopt *ad hoc* policies with regard to which project members have access to ongoing materials created by the project; and finally they are in the early stages of and still unsure about depositing outputs such as data to a common repository or publishing results online (as opposed to via the traditional mechanism of journals).

A third example are the shared 'middleware' software tools, tools which mediate between the shared computing infrastructure and the individual projects or end-user researchers. On this meso-level, we can take case of the Open Middleware Infrastructure Institute (OMII) in the UK as an example. OMII aims to create a suite of widely adopted standard tools which are held and maintained in a repository and made accessible to all researchers. OMII has been promoted by the UK e-Science programme and has also involved the collaboration of commercial software vendors such as Microsoft and IBM. However, despite a push towards open source licensing and standards, the practice has been more complex [9]. It has proved difficult, for example, to achieve standardization, and the efforts to transform the software from the bespoke form that it has within a project into a product that can be shared by many researchers has often proved to be greater than creating the software in the first place.

In short, there have been major efforts to promote the openness of e-Research, and policy debates and debates among researchers [5, 25] continue about the issues that have been described here. It is too early to predict the outcome of a process that is still under way, but by disaggregating 'openness' into a number of levels and component parts, gathering materials on the various levels, and identifying the tensions and forces at work, OeSS will be able to analyze this issue, locate it among the other key themes of the projects (there are clearly connections between openness and, for example, data sharing), and assess the implications for scientific practices.

Open science also has the potential to transform research practices on multiple levels. At one level, having research data openly available enables secondary use which potentially allows less advantaged researchers to participate in scientific discovery even if they don't have access to the resources necessary to gather high-quality datasets. Of course, there are other limitations beyond simply access to data when trying to understand why certain scholars have greater scholarly output [18], but lowering the barrier will potentially increase participation. At another level, open data also exposes research data to the same sort of peer review that scholarly publications now undergo. Currently, there is little provision in the practice of science for datasets to undergo peer review beyond a short methodological description in a publication. If data becomes more open and subject to peer review, the quality of the data being used to create scientific knowledge will potentially be transformed for the better.

2.3.2. Constraints on e-Research

A number of the themes and issues identified by the OeSS project pertain to constraints limiting the uptake of e-Research among scholars. **Privacy**, for instance, is a serious concern to all researchers collecting data on human subjects. While quantitative research data is often relatively easy to anonymize and many quantitative researchers routinely create de-identified or limited data sets, qualitative researchers have not generally had to deal with this issue in the past. Since few qualitative data sets were shared outside the original research team, there was little reason to de-identify the data. The challenge in de-identifying qualitative data is non-trivial compared to the relative ease of simply excluding several columns of quantitative data. Video recordings include images of participants, audio recordings can be identified by the sound of a person's voice, and transcribed interview data often includes references to other people by name and other identifying information. It has been suggested that improving the quality of qualitative data is an important goal for social researchers, and that transparency about the data and the research process are key elements to doing so [6]. The question of how to share such rich data while continuing to protect the identity of research subjects, however, is a major barrier to collaboratively sharing qualitative research data. Whether the promise of gains offered by collaboratively sharing qualitative research data is great enough to overcome these constraints remains to be seen; if it is, it would potentially mark a major transformation in the generally solitary and opaque manner in which much qualitative data is collected.

This also relates to another key theme being developed by OeSS researchers Carusi and Jirotko: **trust**. Trust can take many forms, and includes different sorts of trust, such as the trust people place in artefacts, and the trust people have for those who make use of artefacts [8]. Trust is central to human relationships in general and to scientific collaboration in particular; without trust, people are unlikely to share and collaborate. This is an issue for individual researchers and research participants; on a wider level it is also a question about people's trust in government and researchers generally. Even in Sweden for example, which is generally a 'high trust' country in this respect, to continue to maintain this trust and extend it in the face of new e-Research possibilities that are being developed, takes much effort [4]. While it is therefore possible to build information systems that inspire trust because of high levels of perceived security, that trust can be easily destroyed by the most minor of breaches. Trust between individual scientists is equally hard to maintain since it relies mainly on interpersonal characteristics. The challenge for e-Research, then, is to understand which elements of the research process are most reliant on trust, and to ensure that trust is not undermined by the technologies and processes expected of contributors.

Another concern that has been expressed by e-Research projects pertains to the **economic constraints** academic researchers face when compared to industry players. NCeSS nodes are generally small demonstrator projects with little or no support for the eventual development of mature research tools with the extensive user support that would be required if usage were to become widespread. There is a feeling among some in the e-Research community that the most likely outcome of many of these academic demonstrators will be that industry players will see that

some of them are viable approaches to a problem, and the industry developers will then create a package which will trickle back down to the research community in several years' time [12]. There is already some evidence for this: recent speculation in *Wired Magazine's* blog reports rumours that Google is developing a project (code named Palimpsest) to host open-source science data in a tool that will include commenting and annotating features [16]. Google's resources are far more extensive than those of academic projects; whether this project develops into something more substantial remains to be seen, but it could become a major player in academic research if the success of Google Scholar is any indication. Should well-funded industry players decide to move into the e-Research arena in a serious way, the entire landscape of e-Research may be transformed.

While many other constraints have limited the early uptake of e-Research, these few examples should give an idea of the types of issues we have identified and the potential for transformation in science if these constraints can be overcome. We next turn to several enablers for e-Research.

2.3.3. Enablers for e-Research

Funding shifts

As indicated in the introduction to this paper, shifts in funding in the U.K., the U.S. and elsewhere are major drivers behind the push to e-Research. The shifts at NSF towards transformative research and the growth in the cyberinfrastructure funding program attract new proposals. Likewise, programs in the U.K. such as NCeSS encourage researchers to focus their attention on developing tools and methods that enable e-Research. If these and other programs in the U.S., the U.K., and elsewhere continue to push new funds to e-Research development, the predictions of the funders that e-Research will become a central focus of science becomes a self-fulfilling prophecy, and the likelihood of transformations in the practice of science become much more likely.

New scientific questions

The ability to potentially answer previously unanswerable scientific questions may be the most important driver behind the push to e-Research. This is particularly true in certain scientific fields where access to massive datasets may be the only way to answer particular scientific questions. In biomedical genetics research, for instance, very few of the disorders for which scientists are trying to identify a genetic basis are triggered by a single gene waiting to be discovered. Instead, many disorders are suspected to be triggered by complex multiple gene interactions, and may also be influenced by interactions between genes and an individual's environment. As such, many genetic studies discuss measuring genetic risk for developing a particular disorder, rather than discovering a gene responsible for the disorder. Because this is much more complex than identifying a gene in the population, genetics researchers are finding that they must rely on ever-larger datasets if they are to have any hope of discovering the genetic alleles that place subjects at risk for disease. One way the scientists are doing this by contributing their collections of subject data and DNA to larger collections, such as the GAIN example described above. By increasing the number of DNA samples available at the same time as increasing the amount of data available from each sample, genetics researchers hope to find better indicators of genetic risk in the populations they study.

Another type of research that can only really be done collaboratively is astronomy. Astronomers have long shared resources such as telescopes and sky data. There are currently several international projects designed to create virtual observatories in the U.K. (AstroGrid), the U.S. (NASA's National VO), Europe (Euro-VO), and elsewhere. These projects have been built to enable astronomers to work with data that is stored distributed across the Grid and analysed using parallel processing techniques. The AstroGrid project, for instance, can find, process, and analyse data from a given patch of sky that has been collected from any instrument to which the project has access to the data on the grid. Only the small portions of the data that are the result of calculations need to be transferred back the astronomer, rather than requiring him or her to download a large dataset and work with the data natively. As these tools become more widely used, astronomers are able to spend less time working with data and more time doing analytic science.

A final example of how these collaborative resources can transform practice in a scientific field comes from the field of oceanography. Lamb [15] has described how oceanographers are now able to use remote sensing to transform the practice of event-driven science. The traditional means of collecting oceanographic data was to send ships out to sea and wait for the types of events the scientists needed data on to occur. In recent years, however, oceanographers have been building extensive networks of remote sensing devices located throughout the oceans. Now, the scientists can monitor the readings from their offices. Also, if an interesting event begins about which they would like to collect additional information on-site, they can be notified of this by the remote sensors and then dispatch a ship to sail to the location of interest. This represents an enormous savings in time and money for the oceanographers, and also allows them to collect much more data than before.

If the necessity of harnessing massive resources is a major enabler for e-Research, this may offer a clue to the relatively slow uptake of the collaborative and Grid-based elements of e-Research among social scientists. Social scientists have traditionally been trained as primarily autonomous researchers, working individually or on small teams. The datasets generated are often small and easily managed and analyzed using desktop computing applications. Unlike running complex astronomical calculations, even the largest social surveys rarely tax the processing capabilities of SPSS or SAS on the desktop. For many social scientists, the Grid doesn't appear to offer compellingly attractive applications at this time. Even for social scientists using quantitative statistics or large-scale Social Network Analysis (SNA), their computational requirements do not generally approach those of scientists modelling complex biological, physical or astronomical systems. That may change if some sort of "killer app" is developed, but even so, there will be disciplinary resistance to a change in the normal practices of science.

2.4. Future work

Future work on the OeSS project will move beyond monitoring how various e-Research projects negotiate the constraints with which they are faced -- to include identifying more of the ways in which e-Research realizes the promise of transformation for research, and with what consequences. This will be done through additional case studies, and through additional expansion to a consideration of more international cases, including examination of how national research policies influence the uptake of e-Research techniques.

3. DISCUSSION

While many transformations have taken place in early e-Research projects, a more wide-spread transformation of research practice has remained elusive. The projects have undergone transformation, some notions about the possibility of e-Research have been transformed, and certain scientific fields have seen significant transformations in the way that they work and the types of scientific questions they are able to ask. For many scientists, social scientists and other academic researchers, e-Research is still not central to their work.

One of the challenges in understanding the nature of how e-Research is developing is that many of the projects developing e-Research tools have, until fairly recently, been primarily developing tools that had not yet been released or tested widely in real-world settings. As the efforts have matured, however, it has become feasible to study these projects and to begin developing an understanding of the social shaping that has occurred, and the social, organizational, ethical and legal issues that are associated with the projects. For each of the topics discussed in this paper, the examples are often of projects that have advanced far enough beyond their early stages to have created tools that are being used by scientists. In this way, we are able to examine transformations in tools, projects, scientific practice, and the formulation of scientific questions. In some areas, early results indicate that many transformations have taken place. For instance, at the micro-level, many of the projects have experienced significant transformation from their earliest roots as the initial ideas were shaped by experience. As Schroeder & Fry have observed, "the obstacles to the effectiveness of e-science are not so much technical, as social" [24]. These social barriers are part of the feedback loops that operate as the projects transform over time.

At other levels, however, there is currently less evidence that e-Research tools and e-Infrastructures have yet transformed either meso-level practices of ordinary science or macro-level questions that science is able to imagine and study. One question that remains is whether this lack of fundamental transformation is a temporary problem which will disappear as tools become better developed and more widely used, or whether there are larger issues that make it unlikely that science will be transformed so much as reinforced by e-Research tools. If the former is true, then it becomes just a matter of time and effort to reach the critical mass required for transformation. And if mass adoption of these tools seems an ambitious target, it can be noted that research technologies have, in the past, been drivers for scientific changes because the tools have migrated across different applications and disciplinary domains [23]. If the latter is true, however, then one needs to examine the underlying forces that operate to resist

changes in scientific practice and instead co-opt new tools to the service of existing patterns. Indeed, it is very difficult to predict in advance whether new scientific tools are likelier to support normal science or to enable paradigm shifts, to use the language of Kuhn [14].

We are very early in the story of e-Research. Some elements such as electronically accessible journals are well developed, while others such as Grid-enabled applications for social scientists are in their infancy. The extent to which any of these will prove to be transformative remains to be seen, but we would predict that most will instead tend to reinforce existing patterns of scholarship, particularly in the short term. We will continue to research these issues, however, and eagerly await new developments in e-Research as the projects enabling e-Research mature.

4. ACKNOWLEDGMENTS

This research has been supported by NCeSS and the ESRC through grant RES-149-25-1022. The authors wish to thank NCeSS and ESRC for their continuing support, and also wish to acknowledge the contributions of all the people who have worked on the OeSS project, including Annamaria Carusi, Paul David, Matthijs den Besten, Jenny Fry, Marina Jirotko, and other colleagues at the University of Oxford and elsewhere.

5. REFERENCES

- [1] Copyright Law of the United States of America. 1976. 17 USC 105.
- [2] Copyright, Designs and Patents Act (Great Britain). 1988. Chapter 48 (X) (163).
- [3] Allen, R. 2004. NCeSS: the National Centre for e-Social Science. Available online: <http://www.grid.ac.uk/twiki/bin/view/ELearning/NCeSS>.
- [4] Axelsson, A.-S. and Schroeder, R. 2007. Making it Open and Keeping it Safe: e-Enabled Datasharing in Sweden and Related Issues. In Proceedings of the Third International Conference on e-Social Science, (Ann Arbor, Michigan, USA, October 7-9, 2007).
- [5] Borgman, C. 2007. Scholarship in the Digital Age. MIT Press, Cambridge, MA.
- [6] Brower, R.S. 2000. On improving qualitative methods in public administration research. *Admin. Soc.* 32 (4), 363-397.
- [7] Carusi, A. and Jirotko, M. 2007. From data archive to ethical labyrinth. In Proceedings of the 3rd Annual International e-Social Science Conference, (Ann Arbor, Michigan, USA, October 7-9, 2007).
- [8] Carusi, A., Jirotko, M. and Parker, M. 2006. Trust and Ethics in e-Science Agenda Setting Workshop Report, Oxford e-Research Centre, Oxford. Available online: <http://www.oii.ox.ac.uk/events/details.cfm?id=22>.
- [9] David, P., den Besten, M. and Schroeder, R. 2006. How Open is e-Science? In Proceedings of the 2nd IEEE International Conference on e-Science and Grid Computing, (Amsterdam, The Netherlands, December 4-6, 2006).
- [10] Dutton, W.H. 2007. Reconfiguring Access: A Framework. In Proceedings of the World Wide Science Conference, (Oxford, UK, May 31-June 1, 2007).
- [11] Fry, J., den Besten, M. and Schroeder, R. n.d. Open Science in e-Science: Contingency or Policy? *J. Doc.*

- [12] Halfpenny, P. Comments at NCESS Showcase, Manchester, 2008.
- [13] Hudson-Smith, A. 2007. Digital Urban Blog. Available online: <http://digitalurban.blogspot.com/2007/08/ordnance-survey-and-google-statements.html>.
- [14] Kuhn, T. 1962. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.
- [15] Lamb, R. Keynote speech at IFIP-TC9 World Conference on Human Choice and Computers, Maribor, Slovenia, 2006.
- [16] Madrigal, A. 2008. Wired Science Blog: Google to Host Terabytes of Open-Source Science Data. Available online: <http://blog.wired.com/wiredscience/2008/01/google-to-provi.html>.
- [17] Meyer, E.T. 2008. Forthcoming. Moving from small science to big science: Social and organizational impediments to large scale data sharing. In Jankowski, N. ed. *E-Research: Transformation in Scholarly Practice*, Routledge, New York.
- [18] Meyer, E.T. and Kling, R. 2002. Leveling the playing field, or expanding the bleachers? Socio-Technical Interaction Networks and arXiv.org (Center for Social Informatics Working Paper Series WP-02-10) Center for Social Informatics Working Paper Series, Center for Social Informatics, Bloomington, IN. Available online: <http://www.slis.indiana.edu/CSI/WP/WP02-10B.html>.
- [19] National Centre for e-Social Science. 2007. NCESS webpage, Manchester, UK. Available online: <http://www.ncess.ac.uk/>.
- [20] National Science Foundation. 2007. Proposal and Award Policies and Procedures Guide NSF 08-1, January 2008 (Document number nsf081), NSF.gov, Arlington, Virginia. Available online: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=pa pp.
- [21] Ordnance Survey. 2008. Overview. Available online: <http://www.ordnancesurvey.co.uk/oswebsite/aboutus/yourinfo/ights/>.
- [22] Schroeder, R. 2007. e-Research Infrastructures and Open Science: Towards a New System of Knowledge Production? *Prometheus* 25 (1), 1-17.
- [23] Schroeder, R. 2008. e-Sciences as Research Technologies: Reconfiguring Disciplines, Globalizing Knowledge. *Soc. Sc. Inform.* 47 (2).
- [24] Schroeder, R. and Fry, J. 2007. Social Science Approaches to e-Science: Framing an Agenda. *J. Comput.-Mediat. Comm.* 12 (2), Article 11.
- [25] Willinsky, J. 2005. *The Access Principle*. MIT Press, Cambridge, MA.

FULL PAPER SUBMISSION 2:

One iSchool's Ideas and Identity: Doctoral Training and Research at Rutgers-SCILS 1959-2007

AUTHOR(S):

Mokros, Hartmut B.

ORGANIZATION(S):

Rutgers University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The proposed paper for presentation at the Third iConference addresses the “Nature and Scope of iSchools and iResearch” conference theme. It addresses this theme through report of an in-depth case study of the organizational, social, and intellectual evolution of one early pioneer in the iSchool movement, the Rutgers University School of Communication, Information, and Library Studies (SCILS). The paper first situates SCILS within the iSchool movement, examining its commonalities with and uniqueness from other members of the iSchool Caucus. Thereafter, the paper focuses specifically on the evolution of doctoral research and training, beginning with the program’s development as a disciplinary LIS program, followed by its transformation into an interdisciplinary program with the multidisciplinary merger that founded SCILS in 1982. The paper examines the evolution of ideas developed in dissertation research produced before and after the founding of SCILS interdisciplinary doctoral program in 1987. This includes analysis of the 269 dissertation produced in the program from 1961 to 2007. It compares and contrasts these periods in terms of ideas, methods and the influence played by dissertation supervisors (44 in all). Finally, it examines the influence of key organizational changes on the perceived identity of the school, its relationship with external constituencies, and its intellectual vision into the future.

Of the nineteen academic units who are members of the iSchool Caucus (See Table 1), SCILS is one of seventeen to include “Information” in its name, one of thirteen to be identified as a school within its university, one of five to include “Studies” in its name, and one of four to include “Library” in its name. It is the only member school with three disciplines included in its name and also the only member school within the iSchool Caucus to include “Communication” in its name.

The history of SCILS doctoral program dates to the founding of the Graduate School of Library Service (GSLS) in 1959, with its first doctorates granted in 1961 (See Table 2). In 1978, the GSLS changed its name to the Graduate School of Library and Information Studies (GSLIS). The school’s renaming occurred during a time rife with uncertainty, during which PhD granting library programs throughout the country faced serious institutional scrutiny and threat of closure. The move to delete “Service” and replace it with “Studies” promoted an image of research rather than advanced professional training as the emphasis of its doctoral program, with the adding of “Information” to the school’s name acknowledgment of the every increasing relevance of information and technology for library education and research. Within five years of its renaming, the school was radically transformed through merger. Demoted to departmental status, the GSLIS faculty and master’s program joined SCILS along with the faculty and large undergraduate programs of the Department of Communication and The Department of Journalism and Mass Communication. The GSLIS doctoral program continued in operation until 1987 at which time a new interdisciplinary curriculum and new program name were adopted.

The interdisciplinary program established in 1987 has gone through significant changes in the years since. Most significant has been the program’s reorganization from a five to a three area of specialization model in 1999. This change occurred shortly after two of the school’s three departments modified their departmental names. Having removed “Service” in favor of “Studies” in 1978, prior to the merger with SCILS, LIS now removed “Studies” replacing it with “Science.” The change of name to the Department of Library and Information Science acknowledged the departments increasing intellectual ties with and hires of computer science and engineering faculty. At this same time, with LIS moving closer to the physical sciences, Journalism shifted its future toward the humanities, and away from the social sciences, symbolically expressed through the erasure of “Mass Communication” from its departmental in favor of “Media Studies.”

Although Communication did not pursue a change of name for its department, it greatly focused its curriculum, research and hiring around three areas: social interaction, organizational communication, and mediated communication focusing especially on the development of strength in health communication linked to these areas. The proposal and adoption of a new three area model for the PhD program further supported a rising sense of departmental autonomy in pursuit of research and the training of doctoral students. The new model aligned each of the school’s three departments with one area within the program: Communication Processes, Library and Information Science, and Media Studies.

In the years since the adoption of the three area model has SCILS doctoral program has been remarkable productive in its inclusion of new faculty supervisors and in the number of dissertations completed annual. And yet despite its productivity old tensions have emerged. Those tensions reveal the conundrums faced by a multidisciplinary school whose disciplinary pulls and alternative interdisciplinary partnerships have begun to offer more attraction and reward than another round of collective sense making to secure the commons of an i-identity.

FULL PAPER SUBMISSION 3:

Community Identity: Peer Prestige and Academic Hiring in the iSchools

AUTHOR(S):

Wiggins, Andrea Kathleen

McQuaid, Michael J

Adamic, Lada A

ORGANIZATION(S):

Syracuse University, United States of America

University of Michigan, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Community Identity: Peer Prestige and Academic Hiring in the iSchools

Introduction

Exploring indicators of prestige in hiring networks as they relate to measures of prestige presented in peer rankings provides a new perspective on hiring and identity in the iSchools. This study examines a hiring network for the iSchool community and finds that the perception of prestige among iSchools, as represented by the US News & World Report (USNWR) graduate school rankings, may be improved by hiring choices that strengthen connections within the iSchool community, balanced by increasing the diversity of sources for new faculty.

Prestige rankings are a commonly consulted indicator of image and identity (Gioia and Thomas); for a community in which identity is a matter of concern, providing inclusive prestige scores can position each school in a community context. Prestige ratings based on peer survey responses, published by such groups as USNWR and the National Research Council (NRC), imply a hierarchy of quality and prestige in the institutions reviewed (Morse and Flanigan; Maher et al.). One target audience for the ratings is college-bound students, and as such the ratings project an important aspect of identity with respect to student recruitment. For this reason, it is important to question the value of the survey responses as indicators of academic program quality. In the case of existing rankings such as the prevalent prestige rankings for library and information science (LIS) from USNWR, the community context is incomplete.

Literature Review

Growing interest in the formation of community identity in iSchools inspired conference papers on this theme at the 2005 iConference; some concerns included student recruitment and student placement, which are particularly challenging for a new academic discipline and are critical to the ongoing success of the iSchools (Annabi, Fisher and Mai; Leazer). Identity is a clear root factor in these challenges, as a lack of awareness of the iSchool movement hinders student recruitment efforts, and program graduates must be able to clearly articulate the identity and value of their interdisciplinary studies to secure employment. Further challenges identified at the 2005 iConference pertain to the development of the scholarly community from the perspectives of publication, funding, and interdisciplinary research efforts (Tyworth and Sawyer). In other recently established disciplines such as African-American studies, and professional fields such as MIS, concerns over organizational and disciplinary legitimacy play a strong role in the development of disciplinary identity (Lyytinen and King; Small).

In academia, departmental prestige is often considered a reflection of identity. A variety of studies of academic prestige in the social sciences have shown that departmental prestige is related to faculty hiring practices (Bair; Baldi; Burris; Long, Allison and McGinnis). When hiring is based on some particularistic criteria such as prestige, instead of universalistic criteria, such as scholarly productivity, potentially detrimental effects to the field may result in the form of academic inbreeding and greater stratification of prestige (Bedeian and Feild; Hunt and Blair). Evaluating faculty productivity in the iSchools proves difficult due to their interdisciplinarity, particularly for comparison to prestige (Meho and Spurgin). These prior studies of academic hiring have repeatedly shown PhD program prestige to be much more relevant to post-PhD job placement prestige than scholarly productivity at the time of graduation, and while scholarly productivity has little influence on hiring, hiring has a strong effect on scholarly productivity (Long; Long and McGinnis).

Community Identity: Peer Prestige and Academic Hiring in the iSchools

Methods

This study evaluates whether network measures of centrality can predict the peer survey prestige ratings that are a part of the community context of identity in an academic discipline. The data collected in January 2007 documents the educational pedigrees of 693 full-time faculty members at iSchools. The sampling frame was generated from the faculty listings on iSchool web sites, and the data were collected from iSchool web sites, faculty web sites and CVs, and the UMI Dissertation Abstracts database. In addition, a similar data set of hiring in top-ranked Computer Science (CS) departments was used to compare the findings for iSchools to a more established academic discipline.

The population of iSchool faculty included 674 PhD degrees in 172 distinct programs of study; these programs were coded according to the Classification of Instructional Programs (CIP) and further summarized to form 13 broad disciplinary categories, shown in Figure 1 (Morgan and Hunt). To examine the relationship between hiring and peer prestige, we constructed a hiring network of institutional affiliations by combining ego networks for each iSchool institution. Connections between the schools are based on the institutions from which current iSchool faculty received their PhD degrees, using the university as the unit of analysis. Social network analysis was conducted using GUESS and Pajek software to obtain network measures and generate visualizations (Adar; Batagelj and Mrvar).

Analysis & Discussion

We found that the most common area of study, for 39% of the faculty of iSchools, is Computer & Information Sciences, followed by Library Science for 14% of iSchool faculty members. However, a full 47% of the faculty members studied in other fields, bringing great diversity to the iSchool community. We generated a continuous scale by which to evaluate the interdisciplinary diversity of faculty expertise at iSchools by using an information entropy measure based on the number of faculty with degrees from each disciplinary category (Shannon). This disciplinary diversity measure clearly distinguishes between hiring strategies that cultivate interdisciplinary diversity and strategies that pursue a rich but narrow focus.

The data also provide insight into the practice of self-hiring in the iSchools. Nearly all of the iSchools hire faculty from their own parent institution. There are at least two reasons for this phenomenon; first, the faculty may come from other departments within the institution, and second, the iSchools' hiring choices for faculty specializing in such areas as archives and librarianship are more constrained due to the relative small number of PhD granting programs in these disciplines. The iSchools, on average, hired 13% of their faculty from their own institutions. For the iSchools that had hired faculty with a degree from their parent institution, approximately 64% of the self-hires were graduates of the program that now employs them. In most cases, these are faculty with degrees in library science, supporting the idea that faculty specialization in this area is subject to greater hiring constraint. Self-hiring is not necessarily a case of a school's graduates immediately joining the faculty of the school granting their degrees; it is more likely that a significant proportion of these individuals had their start in academia in another institution and have returned to their alma mater years later as accomplished scholars.

Community Identity: Peer Prestige and Academic Hiring in the iSchools

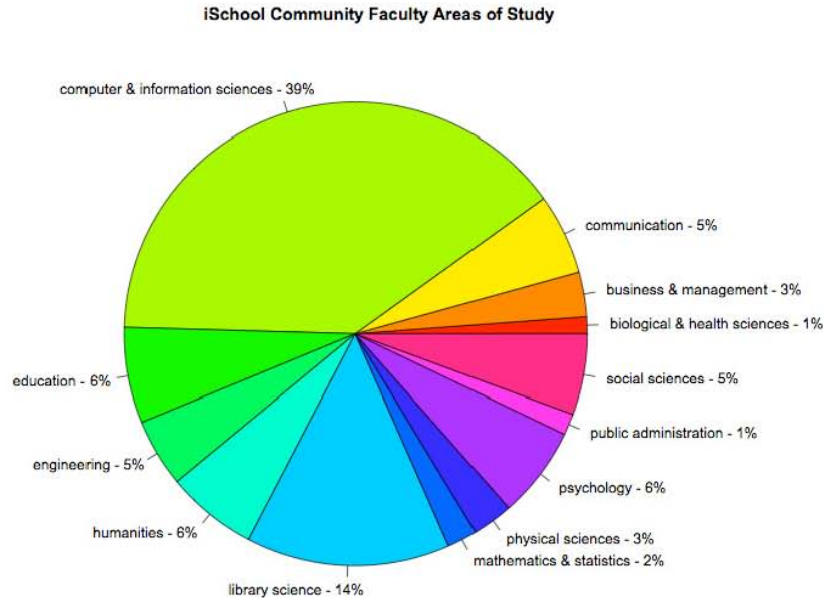


Figure 1: Disciplines represented by the PhD studies of iSchool faculty members.

The weighted, directed network of iSchool faculty hiring, shown in Figure 2, was compared to a similar hiring network for CS. Regression on network prestige and centrality measures was applied to explain the variance in USNWR ratings, and to project inclusive prestige ratings for the full CS and iSchool communities. In the CS network, regression on indegree, weighted PageRank, and betweenness explained 79% of the variance in USNWR ratings with strong significance, $F(3, 22) = 31.7$, $p < 0.0001$. In the iSchools, regression on the number of graduates of each institution who are now on faculty at iSchools, weighted PageRank, betweenness, and hiring diversity explained 77% of the variance in USNWR ratings with $F(4, 6) = 9.3$, $p = 0.01$.

The regression results for these two networks are indicative of underlying similarities in the structure of the two networks, whereas analysis of other aspects of the networks highlights some interesting differences between them, particularly with respect to the diversity of hiring sources accessed by the egos of each network. In the context of the academic communities of computer science and information, the amount of variance explained by regression and level of confidence are evidence that the CS departments form a social structure that is more cohesive and predictable than the iSchool community at this point in time. A much younger discipline, such as the emerging field of information, would not have the same context for describing itself through a peer evaluation as a more established discipline like CS. In the case of the iSchools, these aggregated peer ratings only evaluate a portion of the community on a subset of its programs; however, the regression results indicate that being well connected to the community

Community Identity: Peer Prestige and Academic Hiring in the iSchools

through diverse hiring practices improves the peer perception of prestige for iSchools.

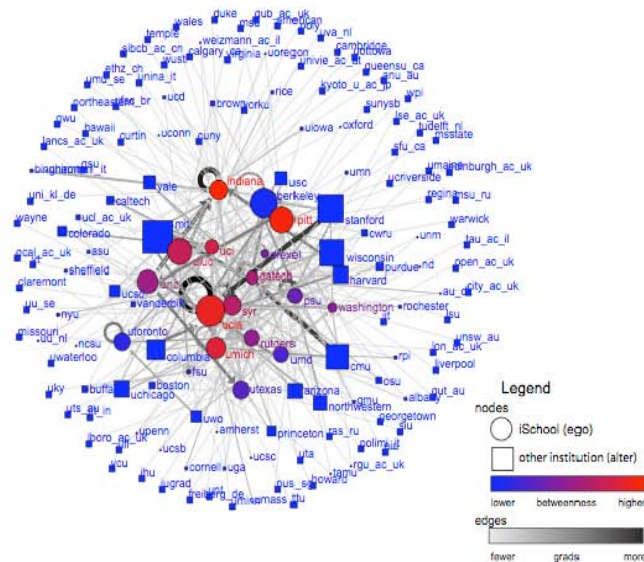


Figure 2: iSchool faculty hiring network visualization, with nodes representing institutions, and edges representing the source of iSchool faculty members' PhDs.

Conclusion

As an artifact of the information school movement, this study holds a mirror up to the iSchool community, but it must be clear that there is no "fairest of them all" despite existing or fitted prestige rankings. The multiplicity of criteria that are relevant to the true measures of success in an institution may be commonly held among many of the schools in the community, but the valuation of those factors is unique to each institutional context.

References

- Adar, Eytan. Guess. Vers. 1.0.1-beta. Computer software, 2006.
- Annabi, Hala, Karen E. Fisher, and Jens-Erik Mai. "Our Academic Life: Challenges Facing I-Schools." *i-Conference 2005*. University of Pittsburgh, 2005.
- Bair, Jeffrey H. "Hiring Practices in Finance Education. Linkages among Top-Ranked Graduate Programs." *American Journal of Economics and Sociology* 62.2 (2003): 429-33.
- Baldi, Stephane. "Prestige Determinants of First Academic Job for New Sociology Ph.D.S 1985-1992." *The Sociological Quarterly* 36.4 (1995): 777-89.
- Batagelj, Vladimir, and Andrej Mrvar. *Pajek*. Vers. 1.11. Computer software, 2006.

Community Identity: Peer Prestige and Academic Hiring in the iSchools

- Bedeian, Arthur G., and Hubert S. Feild. "Academic Stratification in Graduate Management Programs: Departmental Prestige and Faculty Hiring Patterns." *Journal of Management* 6.2 (1980): 99-115.
- Burris, Val. "The Academic Caste System: Prestige Hierarchies in Phd Exchange Networks." *American Sociological Review* 69.2 (2004): 239.
- Gioia, Dennis A., and James B. Thomas. "Identity, Image, and Issue Interpretation: Sensemaking During Strategic Change in Academia." *Administrative Science Quarterly* 41.3 (1996): 370-403.
- Hunt, James G., and John D. Blair. "Content, Process and the Matthew Effect among Management Academics." *Journal of Management* 13.2 (1987): 191-210.
- Leazer, Gregory. "Split Down the Middle, Fuzzy at the Edges: Defining a Field Epiphenomenally." *i-Conference 2005*. University of Pittsburgh, 2005.
- Long, J. Scott. "Productivity and Academic Position in the Scientific Career." *American Sociological Review* 43.6 (1978): 889-908.
- Long, J. Scott, Paul D. Allison, and Robert McGinnis. "Entrance into the Academic Career." *American Sociological Review* 44.5 (1979): 816-30.
- Long, J. Scott, and Robert McGinnis. "Organizational Context and Scientific Productivity." *American Sociological Review* 46.4 (1981): 422-42.
- Lyytinen, Kalle, and John Leslie King. "Nothing at the Center?: Academic Legitimacy in the Information Systems Field." *Journal of the Association for Information Systems* 5.6 (2004): 220-46.
- Maher, Brendan A., et al. *Research-Doctorate Programs in the United States: Continuity and Change*. Ed. Marvin L. Goldberger: National Academies Press, 1995.
- Meho, Lokman I., and Kristina M. Spurgin. "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 (2005): 1314-31.
- Morgan, Robert L., and E. Stephen Hunt. "Classification of Instructional Programs." electronic, 2002.
- Morse, Robert J., and Samuel Flanigan. "America's Best Graduate Schools 2007." 2006.
- Shannon, Claude E. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (1948): 379-423 and 623-56.
- Small, Mario L. "Departmental Conditions and the Emergence of New Disciplines: Two Cases in the Legitimation of African-American Studies." *Theory and Society* 28.5 (1999): 659-07.
- Tyworth, Michael, and Steve Sawyer. "I-Conference Identity and Issues Survey Summary." 2005. 1-6.

Andrea Wiggins
Syracuse University
School of Information Studies
337 Hinds Hall
Syracuse, NY 13244-4100
1-315-558-0122
awiggins@syr.edu

Michael J. McQuaid
University of Michigan
School of Information
1085 S. University Ave
Ann Arbor, MI 48109-1107
1-734-647-9550
mcq@umich.edu

Lada A. Adamic
University of Michigan
School of Information
1085 S. University Ave.
Ann Arbor, MI 48109-1107
1-734-615-2132
ladamic@umich.edu

ABSTRACT

Exploring indicators of prestige in hiring networks as they relate to measures of prestige presented in peer ratings provides a new perspective on hiring and identity in the iSchools. This study examines a hiring network for the iSchool community and finds that the perception of prestige among iSchools, as represented by the US News & World Report (USNWR) graduate school ratings, may be improved by hiring choices that strengthen connections within the iSchool community, balanced by increasing the diversity of sources for new faculty. We compare the academic hiring network for the more established Computer Science discipline to the recently emergent iSchools to explore the relationship between peer prestige and community identity. We also discuss additional observations from the data relating to the interdisciplinary diversity of the iSchool community.

Topics

Nature and scope of iSchools and iResearch

Keywords

iSchools, identity, peer prestige, hiring networks, academic hiring, diversity, social network analysis.

1. INTRODUCTION

Prestige ratings are a commonly consulted indicator of image and identity [1]; for a community in which identity is a matter of concern, inclusive prestige scores can position each school in a community context. Prestige ratings based on peer survey responses, published by such groups as USNWR and the National Research Council (NRC), imply a hierarchy of quality and prestige in the institutions reviewed [2, 3]. One target audience for the ratings is college-bound students, and as such the ratings project an important aspect of identity with respect to student recruitment. This gives us reason to question the value of the survey responses as indicators of academic program quality for the interdisciplinary iSchools, particularly because existing ratings such as the prevalent prestige ratings for library and information science (LIS) from

USNWR do not include the entire community.

We begin with a brief review of the literature related to identity, prestige, and hiring in academia. We then discuss the methods we used to examine the relationship between prestige, as a proxy for identity within a community, and the exchange of social capital represented by academic hiring. We provide a descriptive comparison of the hiring networks for top-ranked Computer Science (CS) departments and the iSchools, and present the results of regression analysis on prestige ratings for each network. We also examine additional aspects of the iSchool community based on the composition of faculty, specifically the phenomenon of self-hiring and the diversity of disciplines represented by the faculty of the iSchools.

2. LITERATURE REVIEW

Growing interest in the formation of community identity in iSchools inspired conference papers on this theme at the 2005 iConference; some concerns included student recruitment and student placement, which are particularly challenging for a new academic discipline and are critical to the ongoing success of the iSchools [4, 5]. Identity is a clear root factor in these challenges, as a lack of awareness of the iSchool movement hinders student recruitment efforts, and program graduates must be able to clearly articulate the identity and value of their interdisciplinary studies to secure employment. Further challenges identified at the 2005 iConference pertain to the development of the scholarly community from the perspectives of publication, funding, and interdisciplinary research efforts [6]. In other recently established disciplines such as African-American studies, and professional fields such as MIS, concerns over organizational and disciplinary legitimacy play a strong role in the development of disciplinary identity [7, 8].

In academia, departmental prestige is often considered a reflection of identity. A variety of studies of academic prestige in the social sciences have shown that departmental prestige is related to faculty hiring practices [9-12]. When hiring is based on criteria such as prestige instead of [more](#) merit-based criteria, such as scholarly productivity, researchers are concerned about potentially detrimental effects to the field in the form of academic inbreeding and greater stratification of prestige [13, 14]. These prior studies of academic hiring have repeatedly shown PhD program prestige to be much more relevant to post-PhD job placement prestige than scholarly productivity at the time of graduation, and while scholarly

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a license for the iSchools to post and disseminate their papers on this site and on a CD. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse this work. The authors are solely

productivity has little influence on hiring, hiring has a strong effect on scholarly productivity [15, 16].

Evaluating faculty productivity for the iSchools proves difficult due to the interdisciplinarity [of the community](#), particularly for comparison to prestige, as there are currently no ratings that are inclusive of the entire community. Although LIS research faculty productivity has previously been measured through publication and citation rates [17], increasing departmental interdisciplinarity and incompleteness of databases poses significant challenges to the validity of LIS faculty productivity studies [18]. In addition, evaluating LIS schools alone would exclude several iSchools that are not accredited by the ALA, and evaluating the iSchools based only on their LIS programs would not appropriately represent the breadth of the relevant faculty expertise. Accounting for the variations across iSchools that is introduced by their interdisciplinarity will remain a challenge in any attempt to rank these schools based on scholarly productivity.

3. METHODS

This study evaluates whether network measures of centrality can predict the peer survey prestige ratings that are a part of the community context of identity in an academic discipline. A network data set representing faculty hiring in the iSchools was generated through manual data collection. While this data would traditionally be collected through a survey of faculty, or from a directory that aggregates faculty survey data by department or academic field, either of these methods would be subject to an unacceptable level of bias due to inaccuracies and omissions. For this reason, the sampling frame was compiled from faculty rosters on institutional web sites, which are updated more frequently than published directories, and are considered the most authoritative public source for this information [17, 19].

3.1 Data Collection

The data set collected in January 2007 documents the educational pedigrees of the full-time professorial faculty members at iSchools. 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 identity that the hiring network seeks to represent. In addition, *Professors emeritae* are more representative of the prior identity states of a school than its current state. We identified full-time professorial faculty by the standard academic titles of *professor*, *associate professor*, *assistant professor*, *associate dean* and *dean*. The data were collected from iSchool web sites, faculty web sites and CVs, and the UMI Dissertation Abstracts database. For each faculty member who could be identified at the time of data collection as meeting the job title criterion, the data collected included their graduate institution and faculty title, the year of their PhD, and the department or school granting the PhD. To address potential validity problems arising from data incompleteness in a relatively small population, the manual data collection took the form of a faculty census with a 100% response rate, yielding 693 terminal degrees held by

687 academics. After adjustments to maintain the PhD degree as the unit of analysis, a total of 674 data points remained.

A similar data set of hiring in top-ranked Computer Science (CS) departments was used to compare the findings for iSchools to a more established academic discipline. Collected in 2005, these data provide the sources of PhD degrees granted to the faculty of 29 computer science and electrical engineering departments, summarizing 1121 faculty PhDs in 527 edges between 123 schools. The departments selected as egos for data collection in this network were the top-ranked 26 programs in the United States and three top Canadian institutions. Reputation survey ratings from USNWR and the NRC were also applied to the CS network data set for analysis of correlations between USNWR ratings and network statistics [2, 20].

3.2 Constructing Hiring Networks

To examine the relationship between hiring and peer prestige, we constructed an iSchool hiring network of institutional affiliations by combining ego networks for each iSchool institution. [Ego networks are constructed based on the set of connections for a focal node](#), called an ego; [each node directly connected to the ego is known as its alter](#).

Connections between the schools are based on the institutions from which current iSchool faculty received their PhD degrees, using the university as the unit of analysis. Each connection between the schools is a weighted, directed link; these links are directed from the graduating institution to the employing institution for each faculty person. The weights for the links represent the number of academics who share these graduation and employment affiliations. Constructing the network unfortunately required merging the two iSchools at Indiana University in order to maintain the institution as the unit of analysis.

Since both the iSchool and CS networks are constructed in the same way, by merging ego networks, they are composed of a set of “inside” nodes for which we have incoming links (information on which other departments they hired from) and the remainder of the nodes for which there are no inbound edges. Those “outside” nodes have only outbound edges, and are included in the dataset if a graduate of the department was hired by one of the departments sampled. In the iSchool network, the inside nodes, or egos, are the iSchools and the outside nodes, or alters, are other institutions that do not have information schools affiliated with the I-School Caucus. In the computer science network, the inside nodes are the most highly ranked departments. This method produces a network with many leaf nodes, [representing those schools](#) that did not provide faculty to more than one inside node, and for which we did not gather information on current faculty.

Both the iSchools and CS departments are portions of the larger academic sphere from which we draw relational information. As ego networks, there is an inherent bias in these data; [while we have complete information about the](#)

[relationships between schools that are egos in the networks, we have incomplete information about those for which we only know that some of their graduates were hired into the academic units that were sampled.](#)

To compare measures of social and network prestige in these networks, hiring the graduate of an institution is considered an endorsement in which patterns of association indicate social exchange.

3.3 Measuring Diversity in Hiring Networks

Schools follow varying strategies to build a strong faculty; some are highly specialized while others are highly interdisciplinary. Two information entropy calculations provide measures of diversity in hiring sources and in areas of subject specialization, by applying the calculation from Shannon [21], $-f \log(f)$, where f is the percentage of the faculty in a given category, either based on their area of expertise or the institution from which they received their degrees. When applied to the hiring data for each school, the hiring diversity measure reflects both the variety and strength of connections to other schools. Schools that hire preferentially from a small handful of highly-respected sources will have low hiring diversity scores and schools that hire from a wide variety of institutions without strong favorites will have high diversity scores. The hiring diversity measure was generated for both networks.

In addition to hiring diversity, an information entropy measure for disciplinary diversity was calculated for the iSchools. The same information entropy formula was applied to the percentage of faculty with degrees in each subject family. The resulting disciplinary diversity scores are highest for the most interdisciplinary schools and lowest for schools with a very strong disciplinary focus, as reflected in the subject areas studied by their faculty.

4. ANALYSIS AND DISCUSSION

The analysis and discussion of the data [are](#) presented in several parts. First, we compare the structures of the iSchool and CS hiring networks, which are similar with regard to their connectedness but demonstrate different hiring tendencies in each community. Next, we examine the relationship between peer prestige and hiring network statistics in both iSchools and CS departments using regression analysis. We also discuss the phenomenon of self-hiring in iSchools, and [finally](#), we discuss the faculty areas of study and the related topic of disciplinary diversity in the iSchools

4.1 Comparing the iSchool and CS Networks

Several global network properties contribute to understanding the context of the interactions that each hiring network represents. The size of the network can be evaluated in several ways; the most apparent measures are the number of nodes and edges, and the ratio of edges to nodes, which gives the average degree of the nodes in the network. The networks statistics shown in Table 1 reveal that the iSchools network has a lower density, lower average degree, lower clustering coefficient, and lower average edge weight than the CS network. The number of degrees summarized in each

network is the primary reason for this difference. While the number of egos in each network plays a significant role in determining these statistics, one notable difference between the two networks is seen in the ratio of alters to egos. The iSchools have more than twice as many alters for every ego as do the CS departments, indicating that the iSchools hire from a greater diversity of sources than the CS departments.

Table 1. Network properties for the CS and iSchool hiring networks.

Network Characteristic	CS Network	iSchools Network
Nodes	123	152
Egos	29	18
Alters	94	134
Ratio of Egos to Alters	3.2	7.4
Edges	572	429
Average Degree	4.7	2.9
Loops	26	17
Total PhD Degrees	1121	674
Average Edge Weight	1.96	1.57
Density	0.038	0.019
Betweenness Centralization	0.21	0.19
Average Distance	2.2	2.3
Diameter	5 (<i>random</i> = 7)	4 (<i>random</i> = 11)
Average Clustering Coefficient	0.23 (<i>random</i> = 0.05)	0.19 (<i>random</i> = 0.08)

Both the number of egos and the average node degree contribute to the difference in link density for the networks; the CS network represents 1121 doctoral degrees with more egos and fewer nodes than the iSchool network, which represents 674 faculty PhDs. The number of edges into which these degrees are summarized provides another point for comparison, shown in Table 1 as the average edge weight for the network, which indicates how strongly the schools in the network are linked. The difference in average edge weights between the networks may represent a number of factors, such as the longer tenure of the CS departments, which has allowed them to build stronger ties within their community over time. These statistics show that the iSchool network is more loosely coupled than the CS network, and continues to demonstrate the difference in hiring diversity between the networks. On average, the iSchools hire from more than twice as many alters as the CS departments.

We also find that both networks exhibit a low diameter and high clustering coefficient, shown in comparison to the statistics for comparable random Erdős-Rényi graphs in Table 1, which are key characteristics of small world networks [22]. Despite other structural differences, the two networks are remarkably alike in these small world characteristics, with very similar average distances, diameters, and average clustering coefficients. Combined with the comparable betweenness centralizations, this suggests that the iSchools network is structurally similar to the CS network in terms of graph connectedness, even though other aspects of the network structures indicate different strategies for hiring.

4.2 Prestige and USNWR Ratings

Regression on network prestige and centrality measures was applied to explain the variance in USNWR ratings. Ratings such as those presented by USNWR and the NRC are considered important as indicators of institutional identity within the larger academic community context. If hiring represents a contributing factor the school's identity, then the centrality measures for the hiring network may explain some of the differences in peer prestige perceptions as reported in the surveys that make up USNWR ratings. In this analysis, the USNWR ratings in LIS were matched to the iSchools for which they were available. Similarly, the USNWR ratings and NRC ratings for the CS departments were collected for the egos of the network.

4.2.1 Peer Prestige Ratings

The USNWR and NRC ratings are based on peer review; both originate from surveys sent to members of the academic community every few years, in which respondents provide ratings of perceived quality for the programs in their discipline. It is reasonable to expect that these data may be confounded by the respondents' preferences for their own alma maters, with the potential effect of inflating the prestige ratings for schools with larger numbers of graduates, simply by virtue of a greater number of their graduates being positioned to respond to the surveys.

The 2006 USNWR ratings used in this analysis were based on a 2005 survey in both CS and LIS, which had respective response rates of 52% and 51%. The USNWR questionnaires for CS were sent to the department heads and directors of graduate studies at sampled institutions. For the LIS survey, questionnaires were sent to deans, program directors, and senior faculty at 50 schools with ALA-accredited master's programs. The NRC and USNWR ratings for CS correlated very strongly, so only the USNWR ratings were used for analysis.

4.2.2 Regression Analysis

In the CS network, regression on indegree, weighted PageRank, and betweenness, explained 79% of the variance in USNWR ratings with strong significance, shown in Table 2. In the iSchools, regression on the number of graduates of each institution who are now on faculty at iSchools (output),

weighted PageRank, betweenness, and hiring diversity explained 77% of the variance in USNWR ratings (Table 3).

Table 2. Regression analysis for the CS Network.

	<i>B</i>	SE <i>B</i>	<i>t</i>
cs-pagerankscore	11.223359	4.294460	2.613 *
cs-betweenness	0.006258	0.000670	9.340 ***
cs-indegree	-0.068210	0.011898	-5.733 ***

* $p < .05$, *** $p < .001$, $R^2 = .8121$, Adj. $R^2 = .7865$,

$F(3,22) = 31.7$ ***

Table 3. Regression analysis for the iSchool Network.

	<i>B</i>	SE <i>B</i>	<i>t</i>
lis-betweenness	-0.004923	0.001131	0.00481 **
lis-pagerankscore	12.604780	2.966607	0.00539 **
lis-output	0.053361	0.010957	0.00279 **
lis-hiringentropy	0.574079	0.247805	0.05972

$p < .1$, ** $p < .01$, $R^2 = .8605$, Adj. $R^2 = .7675$,

$F(4,6) = 9.251$ **

Several network measures were significant in explaining the variance in USNWR ratings, and one variable for each regression proved particularly interesting. The negative coefficient for indegree from the CS regression means that a higher indegree has a negative effect on a school's rating. In effect, the CS departments receive lower ratings if they choose to hire from a greater number of sources. While hiring diversity was rejected as a regression coefficient for the CS network, it was rejected because indegree provided a slightly stronger result, which can be interpreted as evidence of prestige stratification in the network. In contrast, adding hiring diversity to the regression analysis explained an additional 15% of the variance in the iSchool USNWR ratings. Literally interpreted, this means that hiring faculty more evenly from a broader range of schools is a practice that is rewarded with higher prestige ratings in the iSchool network.

The regression results for these two networks reiterate the descriptive comparison of the network characteristics; while there are some structural similarities between them, the iSchools and CS departments differ with respect to the diversity of hiring sources accessed by the egos of each network. In the context of the academic communities of computer science and information, the amount of variance explained by regression and level of confidence are evidence that the hiring in CS departments forms a social exchange

structure that is more cohesive and predictable than the iSchool community at this point in time. A much younger discipline, such as the emerging field of information, would not have the same context for describing itself through a peer evaluation as a more established discipline such as CS. In the case of the iSchools, these aggregated peer ratings only evaluate a portion of the community on a subset of its programs; however, the regression results indicate that being well connected to the community through diverse hiring practices improves the peer perception of prestige for iSchools.

4.3 Self-Hiring

The data also provide insight into the practice of self-hiring in the iSchools. Nearly all of the iSchools hire faculty from their own parent institution. There are at least two reasons for this phenomenon; first, the faculty may come from other departments within the institution, and second, the iSchools' hiring choices for faculty specializing in such areas as archives and librarianship are more constrained due to the relatively small number of PhD granting programs in these disciplines.

In the first case, where faculty are hired from other departments within the institution, the iSchool network departs significantly from the social science departments in Burris' study which hired from their own graduates [11]. Self-hiring in iSchools may in fact represent greater diversity in the interdisciplinary nature of these hires; Pennsylvania State University's iSchool was founded recently enough to have none of its own graduates on faculty, as is also the case for the University of Washington. At PSU, however, nearly 15% of faculty received their degree from PSU, where hiring from other departments in the university may support interdisciplinary diversity within the faculty of the iSchool. In contrast, Washington's faculty is comprised entirely of graduates of other institutions with no self-hires whatsoever, making their iSchool the single exception in the community in this regard.

The iSchools, on average, hired 13% of their faculty from their own institutions. For the iSchools that had hired faculty with a degree from their parent institution, approximately 64% of the self-hires were graduates of the program that now employs them, about 8% of the total population. In most cases, these are faculty with degrees in library science, supporting the idea that faculty specialization in this area is subject to greater hiring constraint. UCLA is an exception in that most of its self-hires were graduates of its education program, rather than library science.

Self-hiring is not necessarily a case of a school's graduates immediately joining the faculty of the school granting their degrees, although [such a](#) scenario has occurred. It is more likely that a significant proportion of these individuals had their first tenure-track employment in academia with another institution and returned to their alma mater years later.

4.4 Faculty Areas of Study

The graduating department or program of study for the faculty of iSchools was a point of interest for two reasons. First, in the event of self-loops, where a university has hired its own graduates, we were interested to know whether these individuals were hired by the same department from which they had graduated, or from a different school within the university. A second reason to examine faculty areas of study is that identity characteristics for each iSchool, such as programs of study and courses, are both influenced by the areas of expertise represented on its faculty, and influential to hiring choices.

4.4.1 Classifying Faculty Degrees

The population of iSchool faculty included 674 PhD degrees in 172 distinct programs of study; these programs were coded according to the Classification of Instructional Programs (CIP) and further summarized to form 13 broad disciplinary categories, shown in Figure 1.

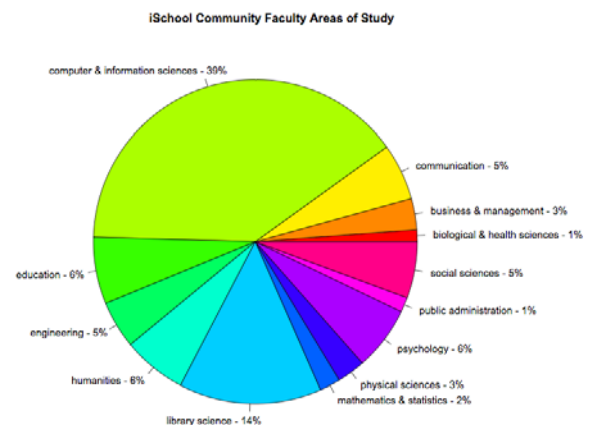


Figure 1. The iSchool community is composed of faculty from a broad range of academic disciplines.

There was some ambiguity regarding how to classify programs entitled *library and information science* or *information and library science*; these were all coded as *library science* because there was a substantial and clearly differentiated population of faculty with degrees in *information science*. The initial coding of the faculty areas of study to CIP families yielded 24 categories; however, some categories such as *family sciences* included very few individuals and other categories, such as *engineering* and *engineering technologies* were sufficiently similar as to provide little additional insight. For analysis purposes, the CIP categories were compressed into the 13 summary categories presented in Figure 1.

The majority of the faculty degrees in the population were in *computer and information sciences*, making up about 43% of the population. The next most common area of study, for 14% of the faculty, was *library science*. Some portion of those degrees classified in the former category might arguably have fit into the latter, if consistent detail about the program of study had been available for faculty with degrees in such areas as

information studies, but data at a level of granularity to allow discrimination between degree programs were not universally available.

As a community, the interdisciplinarity of the field is self-evident, although the iSchools have varying levels of focus on specific aspects of the information field, which seems to be a strategy by which schools differentiate themselves with respect to the community. While permitting the analysis, coding the faculty degree programs and departments into CIP families obscures the true diversity of the academic studies in iSchools, especially within the category of computer and information sciences. The breadth of the academic traditions represented in the schools currently granting degrees in information science or information studies means that the expertise of faculty with degrees in these areas may be very diverse as well.

4.4.2 Disciplinary Diversity in iSchools

While most of the faculty in iSchools studied *computer and information sciences* or *library science*, a full 47% of the faculty members studied in other fields, bringing great diversity to the iSchool community. We generated a continuous scale by which to evaluate the interdisciplinary diversity of faculty expertise at iSchools by using an information entropy measure based on the number of faculty with degrees from each disciplinary category [22]. This disciplinary diversity measure clearly distinguishes between hiring strategies that cultivate interdisciplinary diversity and strategies that pursue a rich but narrow focus. In terms of the diversity of faculty expertise, there is significant variation between schools, as shown in Table 4.

Table 4. Disciplinary diversity is measured using the information entropy function, converted to Z-scores.

iSchools, ($N = 674$)	Disciplinary Diversity
University of Michigan, $n = 39$	1.38
Syracuse University, $n = 33$	1.32
Indiana University, $n = 87$ (both schools together)	1.03
Pennsylvania State University, $n = 48$	0.95
University of Pittsburgh, $n = 31$	0.91
University of California - Los Angeles, $n = 66$	0.67
Rutgers University, $n = 47$	0.67
University of Maryland, $n = 17$	0.55
Florida State University, $n = 25$	-0.01
University of California - Berkeley, $n = 12$	-0.25
University of Washington, $n = 29$	-0.25
University of Illinois - Urbana-Champaign, $n = 22$	-0.31

Drexel University, $n = 24$	-0.32
University of Texas - Austin, $n = 21$	-0.46
University of California - Irvine, $n = 56$	-1.21
Georgia Institute of Technology, $n = 78$	-1.46
University of North Carolina - Chapel Hill, $n = 24$	-1.57
University of Toronto, $n = 15$	-1.66

One interpretation would gauge the interdisciplinarity of the schools by the distribution of faculty representing different areas of study; some schools have chosen to pursue a rich but narrow focus, such as the University of North Carolina, whose faculty's studies are strongly centered around library science and computer and information science. In contrast, schools such as the University of Michigan have made a specific goal of cultivating a broadly interdisciplinary faculty, and have hired academics representing 11 of the 13 aggregated CIP families. The disciplinary diversity measure seems to support this interpretation. Michigan and Syracuse stand out with the highest scores, indicating the greatest interdisciplinarity, while schools such as UNC and the University of Toronto cluster together with the lowest scores, indicating the greatest focus in subject specialization.

The variations in disciplinary diversity indicate different approaches to building an organizational identity through hiring practices at each iSchool, as each faculty represents a different composition of disciplinary expertise. Naturally, a small faculty will tend to represent fewer disciplines, but as Table 4 suggests, size and disciplinary diversity are not strongly correlated ($r = 0.18$). In the iSchools, a full-time faculty of 25 or fewer persons will most likely have faculty expertise in five or fewer broad disciplines; one notable exception is the University of Maryland, where a small faculty of seventeen individuals spanned eleven disciplines. Above the threshold of 25 full-time faculty members, the iSchools usually employ academics with expertise in eight or more academic areas of study.

5. CONCLUSIONS

Finding that peer prestige measures such as USNWR ratings can be predicted with hiring network statistics is reason to question what these ratings really mean to a school's identity. Peer ratings can play an important part in perceptions of a school's prestige and role in the academic community; as these ratings are targeted to prospective students, managing the prestige aspects of image and identity may be a matter of particular interest to iSchool administrators. The iSchool community itself has expressed concern over explaining the academic identity of the information field, a challenge that extends to the degree to which peer prestige ratings do or do not reflect the true community identity. Because the peer prestige ratings are currently subject to accreditation-based

populations for sampling, an interdisciplinary community will continue to face challenges in achieving a useful summary representation of the relative identities of its constituents.

We intend to track the changes to iSchool faculty rosters on an annual basis, generating a series of data sets that reflect snapshots in the evolution of the hiring network structure. This research will monitor whether the interdisciplinary field of information is following the trend of most academic disciplines, in which a stratified prestige structure becomes one of the strongest determinants in the placement of graduates. Comparison to the CS hiring network shows some meaningful differences from this trend, suggesting that the interdisciplinarity of the iSchool community could prevent the level of prestige-based academic inbreeding observed in other disciplines.

There are a number of additional possibilities for future research, such as generating a hiring network of all ALA-accredited institutions for comparison to the iSchools, which might highlight interesting differences between the hiring structures of traditional LIS programs and the interdisciplinary iSchools. In addition, the data from and results of this study could be compared to a complementary network representing iSchool PhD graduate placement. Finally, analysis merging iSchool hiring and PhD graduate placement data sets would offer a more holistic view of the interactions of intellectual exchange in the community.

As an artifact of the information school movement, this study holds a mirror up to the iSchool community, but it must be clear that there is no “fairest of them all” as suggested by prestige ratings. The multiplicity of criteria that are relevant to the true measures of success in an institution may be commonly held among many of the schools in the community, but the valuation of those factors is unique to each institutional context.

6. ACKNOWLEDGMENTS

Special thanks to Dragomir Radev and Sam Pollack at the University of Michigan and Cristian Estan at the University of Wisconsin-Madison for sharing their complimentary computer science department hiring data set.

7. REFERENCES

- [1] Gioia, D., and Thomas, J. 1996. Identity, Image and Issue Interpretation, Sensemaking During Strategic Change in Academia. *Admin. Sci. Quart.* 41, 3, 370-403.
- [2] Morse, R., and Flanigan, S. 2006. America's Best Graduate Schools 2007. U. S. News & World Report online supplement.
- [3] Maher, B., et al. 1995. *Research-Doctorate Programs in the United States: Continuity and Change*. ed. Goldberger, M. National Academies Press.
- [4] Annabi, H., Fisher, K. and Mai, J.-E. 2005. Our Academic Life: Challenges Facing i-Schools. Conference paper. (Pennsylvania State University, 2005). iConference 2005.
- [5] Leazer, G. 2005. Split Down the Middle, Fuzzy at the Edges: Defining a Field Epiphenomenally. Conference paper. (Pennsylvania State University, 2005). iConference 2005.
- [6] Tyworth, M., and Sawyer, S. 2005. I-Conference Identity and Issues Survey Summary. Conference survey results. Retrieved from http://bcs.ist.psu.edu/iconference/notes/iconf_final_results.pdf, October 12, 2006.
- [7] Lyytinen, K. and King, J. L. 2004. Nothing at the Center? : Academic Legitimacy in the Information Systems Field. *J. Assoc. Inform. Syst.* 5, 6, 220-246.
- [8] 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.
- [9] Bair, J. 2003. Hiring Practices in Finance Education. Linkages Among Top-Ranked Graduate Programs. *Am. J. Econ. Sociol.* 62, 2, 429-433.
- [10] Baldi, S. 1995. Prestige Determinants of First Academic Job for New Sociology Ph.D.s 1985-1992. *Sociol. Quart.* 36, 4, 777-789.
- [11] Burris, V. 2004. The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks. *Am. Sociol. Rev.* 69, 2, 229-237.
- [12] Long, J., Allison, P. and McGinnis, R. 1979. Entrance into the Academic Career. *Am. Sociol. Rev.* 44, 5, 816-830.
- [13] Bedeian, A. and Feild, H. 1980. Academic Stratification in Graduate Management Programs: Departmental Prestige and Faculty Hiring Patterns. *J. Manage.* 6, 2, 99-115.
- [14] Hunt, J. and Blair, J. 1987. Content, Process and the Matthew Effect Among Management Academics. *J. Manage.* 13, 2, 191-210.
- [15] Long, J. 1978. Productivity and Academic Position in the Scientific Career. *Am. Sociol. Rev.* 43, 6, 889-908.
- [16] Long, J. and McGinnis, R. 1981. Organizational Context and Scientific Productivity. *Am. Sociol. Rev.* 46, 4, 422-442.
- [17] Adkins, D. and Budd, J. 2007. Scholarly productivity of U.S. LIS faculty. *Libr. Inform. Sci. Res.* 29, 1, 154-158.
- [18] Meho, L. and Spurgin, K. 2005. Ranking the research productivity of library and information science faculty and schools: An evaluation of data sources and research methods. *J. Am. Soc. Inf. Sci. Tec.* 56, 12, 1314-1331.
- [19] Matia, K., et al. 2005. Scaling phenomena in the growth dynamics of scientific output. *J. Am. Soc. Inf. Sci. Tec.* 56, 9, 893-902.

- [20] Maher, B.A., et al. 1995. Research-Doctorate Programs in the United States: Continuity and Change. ed. Goldberger, M. National Academies Press.
- [21] Shannon, C. 1948. A Mathematical Theory of Communication. AT&T Tech. J. 27, 379-423 and 623-656.
- [22] Watts, D., Dodds, P., and Newman, M. 2002. Identity and Search in Social Networks. Science. 296, 1302-130

FULL PAPER SUBMISSION 4:

Technical Authorship: Refiguring the Designer-User Conflict and the Visioning of Collective Technical Futures

AUTHOR(S):

Cohn, Marisa Leavitt

Sim, Susan Elliott

Philip, Kavita

ORGANIZATION(S):

University of California, Irvine, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Technical Authorship: Refiguring the Designer-User Conflict and the Visioning of Collective Technical Futures

Introduction

Authorship is a relatively unexplored metaphor for the practices of technology design despite the presence of narrative techniques in various design methodologies. Narrative and story-telling have been used in software design in such forms as "personas," "scenarios," and "user stories." More fully exploring the metaphor of authorship for the creation of technological systems not only explicates these existing practices but also suggests new approaches to the politics of technology design.

Metaphors for technology design are already acknowledged in existing practices such as Software Architecture, agile software development, and many others. Metaphors reveal and conceal, but are limited in that they cannot provide an exhaustive framework for the evaluation or validation of best practices. Still, as a young discipline, technology design can benefit from a greater diversity of metaphors. Instead of encouraging competition among existing practices, we should perhaps embrace the unique opportunity afforded by this field as a contested domain that is ambivalent in its allegiance to disciplinary boundaries of art, science, and social science. Thus, even as the field seeks to produce long-lasting knowledge, we must be wary of eschewing emerging practices that do not fit existing paradigms. Authorship, used as a lens to examine software design practices, not only reveals new information about older practices but also defines a new space for understanding emerging practices that have been thus far set aside by academia as counter-productive.

This paper introduces the metaphor of "Technical Authorship." Technical Authorship is a way of looking at technology as not an artifact that is constructed, but as a vision of the future that is authored. One of the strengths of this metaphor is it opens the door to collaborative authoring, that is, the creation of shared technical futures by technical experts and non-technical non-experts. This metaphor highlights two overlooked aspects of approaches to technical systems design: 1) that the divide between the "designer" and "user" is a conflict of contested authorship; and 2) that design methodologies contain rhetorics of authorship which designate certain understandings of the "designer" and the "user."

This paper is part of a group three related papers being submitted from the Department of Informatics at University of California, Irvine. These papers address the theme of bringing the promise of IT to under-addressed groups. This paper considers the challenging of making the creation of technical futures more inclusive and more egalitarian between the "technical" and the "non-technical." A second paper examines the participation of software developers from India in a global community of practice. The third and final paper considers the problem of bringing mobile computing to infrastructure-poor regions, such as South Africa. In addition to their thematic commonality, these papers use a broader analytical lens than is typically applied in software engineering and human-computer interaction, and includes social and historical considerations.

Methodologies as Rhetorics of Authorship

While many design approaches have addressed the divide between the designer and the user, these do not view the conflict as one of contested authorship. The divide between the designer's vision and user's experience is most frequently addressed because of the urgent and abrupt way that the latter can rupture the former. Massive failures of systems are often cited as evidence that design should "involve the user." Through the lens of Authorship it becomes clear that both expert and non-technical envisionings of technology are forms of authorship and that the design approach is often a method for negotiating co-authorship.

Many approaches to technical systems design encourage co-authorship by involving the "user" as a participant in earlier stages of the design process. Methods such as Participatory Design seek a political aim to de-instrumentalize the user by making the user the subject rather than the object of the design process. However, no matter how much the "user" is empowered through participation in the design process, the role of the designer as a removed, "modest witness" [7] of "use" is never fully extricated. Technical Authorship helps to reveal this tension that the "user," even when reconfigured as actor or participant, is the central figure of all design practices.

All methodologies of design must address the role of the designer, the role of the user, and the relationship between them. When we examine literature on methodologies as rhetorical, we can derive an

understanding of the “designer” and “user” figures. While many design methodologies reveal the designer as author-subject and the user as authored-object, other methodologies reveal the designer as the object of the design process as well. Considering the lens of Authorship enables us to add a new dimensionality to the considerations of design methods. Both designer and user can be either subject or object, along a continuum. Designer and user as categories can be redefined as the technical expert and the non-technical non-expert who are divided by technical language.

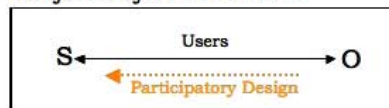
Figuration

We use two analytical devices in our critique, the “figure” and the “modest witness.” Haraway’s concept of the figure is one who “collects up the people; a figure embodies shared meanings in stories that inhabit their audiences” (23). Creating a figure of the “user” as it is understood in various technological design methodologies allows us to see how the role is situated on a subject-object continuum. In addition, we use Haraway’s idea of the modest witness as “the legitimate and authorized ventriloquist for the object world, adding nothing from his mere opinions, from his biasing embodiment” (24). In some design methods, the “designer” is imagined as a kind of modest witness, set apart through a kind of special objectivity. However, our critique reveals that the designer can also be an *immodest* witness.

The user-figure: subject or object?

Literature on Participatory Design, contextual inquiry, “situated actions” and the use of Personas, scenarios, and user stories within design practice, are examined in this paper to arrive at the current figure of the “user.” Each of these techniques addresses the figuring of the “user” as a subject or an object in the design process. Personas are frequently used in large-scale systems design. Companies conduct a research phase during which social science techniques are used to study the actual users and derive data in the form of archetypal characters. Participatory Design explicitly critiques these kinds of methods, which reify the user-object and so seeks to re-engage with the user as a subject. This kind of movement can be illustrated as one along a continuum of the user as subject or object, pictured below. The user can never be fully realized as a subject, since the user is a “figure” in the rhetoric of design and so belongs to this narrative.

Subject-Object Continuum

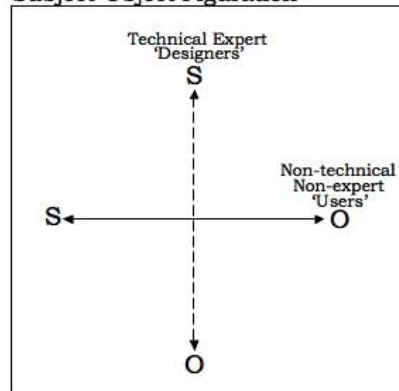


The designer-figure: modest witness or hero's journey?

This same literature can provide an understanding of the “designer” figure. Many practices provide techniques by which the designer can arrive at an objective understanding of “use.” Contextual inquiry warns that this is not entirely possible and focuses on the agency of the user in defining use. Still, even these techniques are used to help the designer witness and provide an account of a kind of objective “use.” Less worried about the reification of the user, personas and scenarios are said to help the designer step out of and transcend his role in as if on a hero’s journey to discover true “use.” Even when expressly conscious of the reification of the “user,” as in Participatory Design, the designer figure himself presents a significant problem to the rhetoric because he embodies the power-knowledge of the technical language.

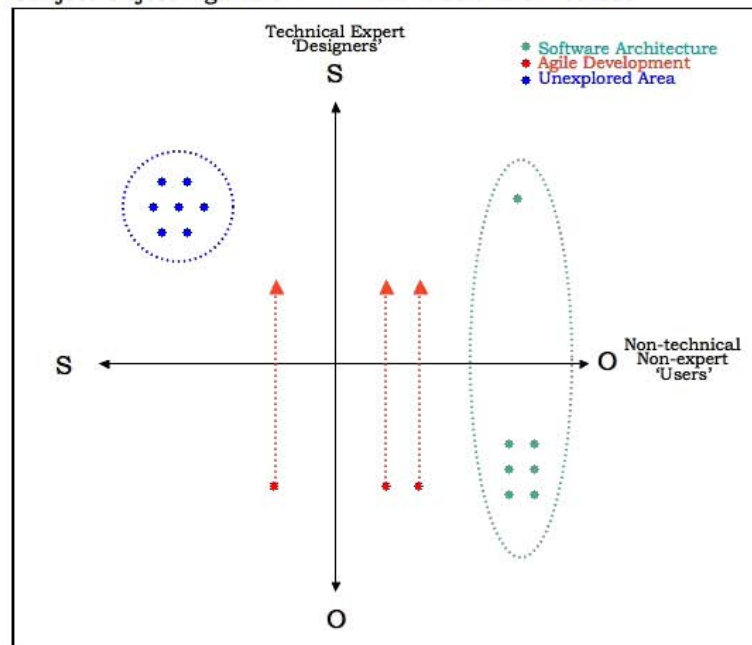
This leads to adding the designer into the Subject-Object Continuum as pictured below.

Subject-Object Figuration



This reveals that the designer can also be reified in the rhetoric of technical authorship. Coders can be seen as the instrument for obtaining technological artifacts as much as users can be seen as the instrument for obtaining objective "use." This new dimensionality provides a space for understanding practices such as Agile Software development, which has otherwise been seen to contribute little to the field.

Subject-Object Figuration in Different Software Methods



The division of those involved in design into the categories of technical expert and non-technical non-expert demonstrates that the user-figure and designer-figure can be located in diverse relationships within this subject-object continuum. Different tensions are revealed by locating methodologies within this continuum. Furthermore we note the relatively unexplored are where both designers and users are situated as subjects.

This reframing of technology design as Technical Authorship leads to the breakdown of the designer

into the roles of expert, author, and especially translator between the technical and non-technical or popular languages. It lays out a different set of obligations and accountabilities for designers of technology. Not only can we work to make the “user” and the “designer” subjects in design practice, technical experts could succeed in bridging the gap between designers and users by defining a new obligation to translate technical languages. Only when we collapse these categories of technical expert and non-technical non-expert might it truly be possible to envision collective technical futures.

References

1. BECK, K. 2000. *Extreme Programming Explained: Embrace Change*. Addison-Wesley Professional.
2. CARROLL, J.M. 2000. Five reasons for scenario-based design. *Interacting with Computers* 13, 43-60.
3. ERICKSON, T. 1996. Design as storytelling. *interactions* 3, 4 (Jul. 1996), 30-35. DOI= doi.acm.org—234813.234817
4. GRUDIN, J. 2003. The West Wing: Fiction Can Serve Politics. *Scandinavian Journal of Information Systems* 15, 73-77.
5. GRUDIN, J. AND PRUITT, J. 2002. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. *Proc.PDC* 144-161.
6. GULLIKSEN, J., GÖRANSSON, B., BOIVIE, I., BLOMKVIST, S., PERSSON, J. AND CAJANDER, Å. 2003. Key principles for user-centred systems design. *Behaviour & Information Technology* 22, 397-409.
7. HARAWAY, D.J. 1997. *Modest-Witness@ second-Millennium. Femaleman-Meets-Oncomouse: Feminism and Technoscience*. ROUTLEDGE CHAPMAN & HALL.
8. HOLTZBLATT, K. 2002. Personas and contextual design. Retrieved August 17, 2003.
9. JOHANSSON, M. AND MESSETER, J. 2005. Present-ing the user: constructing the persona. *Digital Creativity* 16, 231-243.
10. JOHNS, A. 2002. The Ambivalence of Authorship in Early Modern Natural Philosophy. In *Scientific Authorship: Credit and Intellectual Property in Science*, M. BIAGIOLI AND P. GALISON, Eds. Routledge, UK, (Chap 3) 67-90.
11. KYNG, M. 1998. Users and computers: a contextual approach to design of computer artifacts. *Scandinavian Journal of Information Systems* 10, 7-43.
12. LATOUR, B. 1994. *Science in action: how to follow scientists and engineers through society*. Harvard University Press, Cambridge, MA.
13. MADSEN, K. H. AND AIKEN, P. H. 1993. Experiences using cooperative interactive storyboard prototyping. *Commun. ACM* 36, 6 (Jun. 1993), 57-64. DOI= doi.acm.org—153571.163268
14. MULLER, M. J. AND KUHN, S. 1993. Participatory design. *Commun. ACM* 36, 6 (Jun. 1993), 24-28. DOI= doi.acm.org—153571.255960
15. ROGERS, Y. 1994. Exploring obstacles: integrating CSCW in evolving organisations. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (Chapel Hill, North Carolina, United States, October 22 - 26, 1994). CSCW '94. ACM Press, New York, NY, 67-77. DOI= http://doi.acm.org/10.1145/192844.192875
16. SHAPIN, S. 1994. *A Social History of Truth: Civility and Science in Seventeenth-century England*. University of Chicago Press, Chicago.
17. SUCHMAN, L.A. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press,
18. Vredenburg, K., Mao, J., Smith, P. W., and Carey, T. 2002. A survey of user-centered design practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves* (Minneapolis, Minnesota, USA, April 20 - 25, 2002). CHI '02. ACM Press, New York, NY, 471-478. DOI= doi.acm.org—503376.503460

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

Technical Authorship: Refiguring the Designer-User Conflict and the Visioning of Collective Technical Futures

Marisa Leavitt Cohn
Department of Informatics
University of California, Irvine

Irvine, CA 92697-3440
+1 949 824 4047

mlcohn@ics.uci.edu

Susan Elliott Sim
 Department of Informatics
 University of California, Irvine
 Irvine, CA 92697-3440
 +1 949 823 2373
 ses@ics.uci.edu

Kavita Philip
 Department of Women's Studies
 University of California, Irvine
 Irvine, CA 92697-3440
 +1 949 824 7092
 kphilip@uci.edu

ABSTRACT

This paper argues for the application of authorship analysis to technology design. It extends techniques used in Science Studies to investigate scientific authorship in order to define a concept of technical authorship. To illustrate the potential of this approach, authorship analysis is applied to particular prescriptive software design methodologies including Participatory Design and Agile Software Development. The results of this analysis are the recognition that: a) design methodologies are rhetorics of authorship, b) the designer-user relationship can be seen as a conflict of contested authorship, and c) this conflict can be seen within the tensions in the figuration of the user and the designer along a subject-object continuum. "Figuration," a technique used by Donna Haraway, brings about a new understanding of the centrality of the user in the design process as well as the role of the designer as the contested author technical futures. The implications suggest that design researchers might use authorship as a new approach to the politics of design by re-figuring the designer and user rather than collapsing the distinction between these roles.

Topics

Information organization
 Information technology and services

Keywords

Authorship, Science Studies, Design Methodologies, Figuration, Politics of Design.

8. INTRODUCING AUTHORSHIP

Authorship is a central subject of study in cultural and literary studies and has received special attention in Science Studies. Michel Foucault first drew attention to the social construction of the concept of authorship, its historical origins and the functions that it serves in his essay "What is an author?" (1977) [9]. He points to the fact that there is both an empirical individual author, and an "author-function" that requires the construction of a figure of an author "who is

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse

outside and precedes" the authored text. The author-function and its figure arose as a new form during the 18th century conditioned by a growing culture of private property, [5 (author citing Foucault)].

Donna Haraway and others in Science Studies have developed this approach further to analyze the scientific author, its origins and implications [3, 7, 17, 19]. The scientific author is the authoritative voice of objective reasoning, that produces facts out of the experimental practices of the sciences. Latour [22], Shapin [29], Haraway [15], and others have helped to explain the origins of how this particular authorial voice came to be as well as it how it continues to be taken for granted. Latour [22] especially focuses on how the everyday practices of the laboratory aid in the social construction of facts, thus unveiling the scientific author (empirical individual) of his garb of scientific objectivity.

Haraway's focus shifts away from unmasking the scientist and towards the emergent author-figure who she calls the "modest witness" (borrowing the term from Shapin and Schaffer, *Leviathan and the Air-Pump*, 1985). The "modest witness" is "the witness whose accounts mirror reality" through "self-invisibility." This modest voice is that which "pays off its practitioners in the coin of epistemological and social power." It is "the virtue that guarantees that the [scientist] is the legitimate and authorized ventriloquist for the object world, adding nothing from his mere opinions..." [15].

This scientific authorial voice, which extends to much of academic writing in the social sciences and technical fields as well, is one that puts the objects of study in supposedly plain view while the author as a subject disappears from the text. The "modest witness" is used by the scientific author (the empirical individual) to offset his role as a subject. Authorship in the sciences appears subject-less. It is a world of objects that speak for themselves. This lack of a subject we now take for granted as that which lends authority to the scientific text, but this authority had to be shaped through the creation of the figure of the "modest witness" who steps aside, and out of view.

In this paper we will extend analysis of authorship to the field of technology design relying primarily on the techniques used by Donna Haraway in her book: *Modest_Witness@Second_Millennium.FemaleMan@_Meets_OncoMouse™*. One of Haraway's techniques is "figuration" which she uses to draw up this character of the "modest witness."

Haraway's concept of the figure is one who "collects up the people; a figure embodies shared meanings in stories that inhabit their audiences." Figuration is difficult to explain but we are familiar with figures, in particular the idea of the "Christ-figure" where certain emblematic features of a character come together through the many metaphoric uses of it in different contexts or stories.

The figure in technology design that readily emerges through authorship analysis is that of the "user." The user is central to technology, computational systems, and product design. In sociological studies of technology design, however, the word has been taken up as highly problematic [13]. It is flagged in various debates within technology design as in the words "user-friendly," "user-centered," and "user experience" (aka UX). It is critiqued, ironically, for its lack of utility since it provides no context, characteristics, or sense of expertise. In fact "UX" responds to the last challenge by establishing "experience" as the user's area of expertise. "User" limits agency on the receiving end of technology to a single action: use, and so has been re-conceived as actor, knowledge-maker, or participant. "User" can provide only a generic sense of what is user-friendly or usable. The design researcher's response to that might be to use methods of "contextual inquiry" or "situated actions" [30]. Grudin has pointed out that the word "end-user" is a retronym that tells us more about the history of the man-machine relationship than about the real empirical user.

Yet, there is an apparent difficulty in dispensing with the word entirely. Authorship is a lens that can help to explain this persistence of the user-figure as problematic rather than trying to explain the particular problems it causes. Haraway's techniques for authorship analysis, using figuration, suggest that what is most interesting about the user is the way it is "figured" through technology design.

The other figure that emerges from authorship analysis of technology design is that of the designer. It is not strange that we as designers and design researchers understand this figure far less than we do the user-figure. Unlike the modest witness, the designer-figure is still very much in flux much as the scientific author-figure was in the 18th century. The many competing design methodologies point to an unclear author-figure and even an ambivalence towards the designer as author.

In fact, authorship analysis suggests that the politics of technology design exist in a conflict between the designer and the user that is one of contested authorship. Most often is the designers themselves that contest their own role as author and initiate methods to involve the users as authors in the design process. But there are also cases where software developers or engineers wish to empower themselves as authors of the code in cases where they feel they have become instrumentalized by a design methodology.

9. MOTIVATIONS

There are a number of motivations for applying authorship

analysis to technology design. First, there are many narrative techniques used in various design methodologies such as the use of "personas" [6, 11, 12, 16, 18], "scenarios" [4], "user stories" [1], and "storyboarding" [8, 24] in software systems development. These methods have been shown to have communicative power in the exchange between designers of technology and their client or user group, but these remain isolated methods that have not been theorized about as a whole.

Second, in technology design it is openly acknowledged that approaches to design can be metaphoric. Software development is likened to engineering or architecture and design methodologies such as 'waterfall,' 'agile,' and 'spiral' each make use of a metaphor whether rigorously or no. Metaphors both reveal and conceal, and are limited in that they cannot provide an exhaustive framework for the evaluation or validation of best practices of design. Still, as young disciplines, technology design and design studies can benefit from a greater diversity of metaphors. Rather than encouraging a winner-takes-all atmosphere among competing methodologies, we should take the time to investigate many metaphors for design, especially since the field is still a contested domain with allegiances that cross disciplinary boundaries of art, science, and social science. As an emergent field seeking growth and foundational knowledge, we must be wary of eschewing practices that do not fit existing paradigms. Authorship as a lens for analysis, not only reveals new information about older design practices, but also defines a new space for understanding emerging practices that have thus far been set aside by academia as counter-productive.

Third, technical authorship is a metaphor that opens the door to collaborative authoring, that is, the creation of co-authored technical futures. With many technology design studies scholars interested in the politics of design, authorship can shed new light on the power struggles within design and suggest new paths to empowerment. Additionally, Haraway's method of figuration causes a different sort of reflection on design practice than other approaches to the politics of design. Figuration allows us to think about how we might want to shape (or "refigure") the role of the user and designer instead of resisting this dichotomy. (Haraway suggests a method of "diffraction" for refiguration, which we will not go into in this paper.)

And finally, there are many parallels between Science Studies and Technology Studies that are only beginning to be explored. In this paper we will be drawing some parallels (that at times are a bit shaky) between technical authorship and scientific authorship with the hope that future work that takes up this approach might lead us to a better understanding of the relationship between science and technology.

10. TECHNICAL AUTHORSHIP

The concept of authorship in the sciences, which was introduced above, is that of an author who presents facts about the natural world he investigates. But authorship of technology seems to differ from scientific authorship in that

it is less a particular voice found in the texts designers produce than the very material shape of their technologies. In Science Studies, however, much attention has been drawn to the way that texts of all kinds, from reports, to talks, to notebook scribbles [28] and the scientific instruments play a role in authoring 'science' [15]. In design it is often 'use' itself that is authored, where the technological product, the design methodology, and even the users themselves, provide the assemblage that shapes 'use.' So the material nature of authoring technology is not a major issue.

However, in order to extend authorship analysis to technology design the idea of technical authorship still needs to be introduced in order to clarify how authorship applies to design of technology. Technical authorship is a way to look at technology not as only an artifact that is constructed, but also as a vision of the future that is authored. It is a technical future that is authored through the practices of the designer, which can include any number of activities such as requirements engineering, prototyping, iterative development, and especially through its successful implementation and use. Authorship is just one lens or metaphor through which to examine these design practices, but it is one that is relatively unexplored despite the presence of narrative techniques in various design methodologies. This paper will apply analysis of authorship to some design methodologies, in order to reveal just a sampling of the potential outcomes that this approach can bring. In particular, it is the opinion of this paper's authors, that certain practices in design that are often considered a-theoretical or a-methodological may be better elucidated through this alternate metaphor of technical authorship.

11. AUTHORSHIP ANALYSIS OF TECHNOLOGY DESIGN

11.1 A Note on Scope

It would be impossible to exhaust the potential for analyzing authorship in technology design in a single paper. Instead we will provide a single attempt to look at software design methodologies through authorship analysis as a way to demonstrate the powers of this approach. In our research lab's ongoing research we contend that design methodologies exist in two forms, prescriptive and empirical, and it is infrequent that a prescriptive method is followed precisely or that empirical design is ever a-methodical even when it does not follow a prescribed method [10].

In this paper we will attempt to investigate the rhetorics of authorship located in prescriptive design methodologies, but we also want to point out that there would be equal if not greater value in applying authorship analysis to investigate empirical design practices where a specific technical future (use/product) is authored. By approaching design methodologies as rhetorical devices we will also touch upon the authorial voice of those design researchers who are writing the design methodologies. You will notice in the quotations provided that it wanders a great deal. At times it is the scientific author, modest, presenting design as a scientific object of study. At other times it is a voice that resembles

inspirational or self-help writing, leading the reader/designer through personal development to better design practice.

11.2 Design Methodologies as Rhetorics of Authorship

The first thing that becomes clear in looking at prescriptive design methodology literature is that nearly all design methodologies must address the relationship between a user and a designer (though there is some design literature which treats design much more as an artistic and craft approach and disregards the user and 'use' entirely). However, while there are many design approaches that explicitly address the divide between the designer and the user, these do not view the conflict as one of contested authorship. Rather, the divide between the designer's "vision" and the user's "experience" is most frequently addressed because of the urgent and abrupt way that the latter can rupture the former. Massive failures of systems are often cited as evidence that design should "involve the user." Through the lens of authorship it becomes clear that both technical expert (designer) and non-technical non-expert (the user) are envisioning the technology in different ways. The design process then appears as a negotiation or conflict over authorship, which in many cases does not truly resolve.

Many approaches to technical systems design seek to involve the user explicitly in the design process. Methods such as Participatory Design [14, 21, 25] seeks a political aim to de-instrumentalize the user who can often be the subject of power alliances between designers and clients. Participatory Design has revealed design as embedded within a politics of the workplace. Because of this it has focused primarily on the conflict between management and the worker and on combating the ways that technology can instrumentalize the user. User-centered design, on the other hand, has focused on the conflict as a research problem and involves the user for a clearer "understanding of user and task requirements," [31]. This approach may produce technology that does not instrumentalize the user in a Heideggerian sense, but does forge the user into the designer's instrument for understanding use.

In Agile Software Development and Extreme Programming the tension shifts in the other direction. The technical engineer is seen to have been disempowered in the workplace, unable to take pride in the craft of software design, and instrumentalized as the tool through which software is enacted. The "Agile Manifesto" [2] and books like "Extreme Programming Explained: Embrace Change," by Kent Beck [1] introduce a message to empower the programmer or software engineer through a kind of bottom-up, subversive self-management technique.

Through the lens of authorship both Participatory Design and Agile Methods can be seen as attempts at co-authorship, though this analysis will shift the metaphor away from that of democratic participation or involvement used by these design researchers. The impact of this shift will be a focus on the tension between author (subject) and authored (object). This turns out to be extremely useful in understanding emerging software design methodologies that have so far been under-theorized.

11.3 The Current State of (Contested) Authorship in Technology Design

A literature review was conducted on prescriptive software design methodologies including Participatory Design [14, 21, 25], contextual inquiry [16], “situated actions” [30] and the use of Personas [6, 11, 12, 16, 18], scenarios [4], and user stories [1]. Authorship analysis, and use of Haraway’s figuration technique, resulted in the recognition that authorship is currently contested in technology design. The results of this analysis are summarized here by illustrating how figuration can fall along a continuum of subject to object.

With the modest witness we see an authoritative author-figure who is at the same time made subject-less. In the rhetorics of technology design there is a tension where the user-figure can either be a subject or an object.

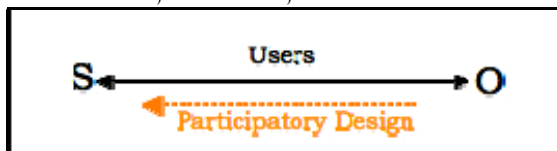


Figure 1. The Subject-Object Continuum

Personas are a technique used frequently in large-scale systems design. Companies conduct a research phase during which social science practitioners study the actual users and derive data in the form of archetypal characters. By the end of the research phase, users (empirical individuals) are transformed into objects; they are in a sense authored by the researcher. Participatory Design, on the other hand, explicitly critiques these kinds of methods, which reify the user, and seeks to re-engage with the user as a subject. This kind of movement can be illustrated as one along a continuum of the user as subject or object (Figure 1). The user can never be fully realized as a subject, since the user is a “figure” in the rhetoric of design and so belongs to its narrative.

Technology design rhetorics reveal a corresponding tension within the “designer” figure. Many design methodologies outline techniques by which the designer can arrive at an objective understanding of “use.” Contextual inquiry warns that this is not entirely possible and focuses on the agency of the user in defining use. Still, even these techniques are used to help the designer witness and provide an account of a kind of objective “use” [16]. Less worried about the reification of the user, personas and scenarios are said to help the designer step out of and transcend his role in as if on a hero’s journey to discover true “use.” Even when expressly conscious of the reification of the “user,” as in Participatory Design, the designer figure himself presents a significant problem to the rhetoric because he embodies the power-knowledge of the technical language.

Authorship analysis suggests that we add the designer into the Subject-Object Continuum as pictured below (Figure 2).

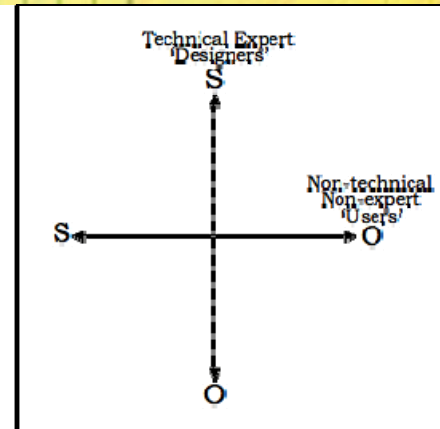


Figure 2. Figuration of the Designer and the User

This reveals that the designer can also be reified in the rhetoric of technical authorship. Coders can be seen as the instrument for obtaining technological artifacts as much as users can be seen as the instrument for obtaining objective “use.” This new dimensionality provides a space for understanding practices such as Agile Software development, which has otherwise been seen to contribute little to the field.

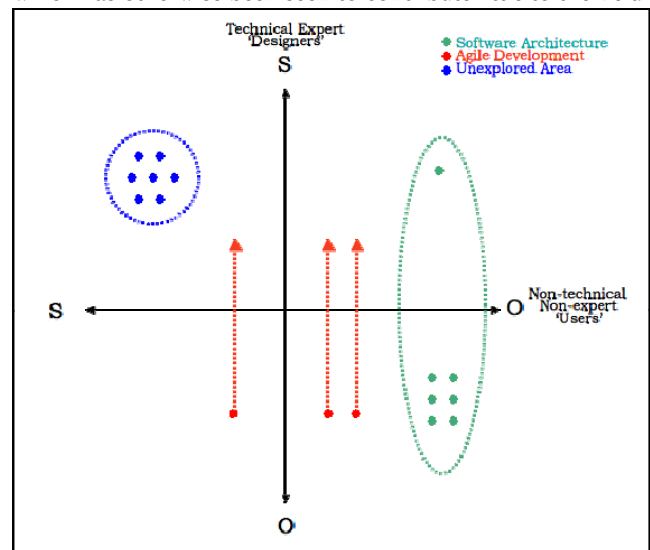


Figure 3. Figuration in Different Software Methods

Agile can be seen here (Figure 3) as a political movement to engage the software developer as subject instead of object. The division of those involved in design into the categories of technical expert and non-technical non-expert demonstrates that the user-figure and designer-figure can be located in diverse relationships within this subject-object continuum. Different movements are illustrated by locating methodologies within this continuum. Furthermore we note the relatively unexplored area where both designers and users are situated as subjects.

These tensions between subject and object and the current state of affairs in the rhetorics of technical authorship suggest that the role of the designer as author is contingent on the

figuration of the user and likewise that the role of the user as author is contingent on the figuration of the designer. Before we attempt to address these tensions politically through projects to explore the space where both designer and user are subjects, we should first aim to understand how these two are figured in design methods and practices.

11.4 Figuration of the user and designer

We will now turn to the rhetorics of authorship found in design methodology literature and begin to shape the figure of the user and the designer that we found through our literature search. When viewed as methodologies, each approach may be seen to align with or critique another approach. But with the technique of figuration, what we look for is the emergent figure of the user, not something generalized from the literature.

The user-figure that emerges is unwieldy, difficult to communicate with, and repellent. In the everyday practice of software engineers it is a "common sense" notion that engineers do not want to speak with actual users, that they are incapable of communicating with users, and that users are irksome in their language and demands. Latour [22] refers to the "fierceness" of disputes in the early stages before a fact becomes a fact. There is a similar fierceness in the early stages of design when requirements are not fleshed out, and this antagonism arises between the user and the designer.

The user-figure is a point of access to a naturalized "use."

Designers who advocate for user-centered design, or participatory design, will criticize projects that do not involve the "real user" in the design process saying that the wants and needs of "real users" can never be known without speaking directly to users.

In large-scale projects designers will often be speak to users, without such political motivations, as a pragmatic way to understand the technology's future use. In this case a researcher will often serve as a proxy to study, interview, and observe users and communicate "use" back to the designer. In this case the user-figure is a way to ward off false use. He is a test for the purity of use.

This purity is set against the impurity of the designer, as a solution to the issue of communication. The trouble with the engineer is that he is "bad at" talking to users, and there is also the issue of scale when the user group and the company hired to design technology may both be too large to interact in a personal way.

The designer-figure must have access to the user-figure because "Getting it right" matters, particularly in large projects where massive failures can arise. There are many invaluable contributions of research in this area highlighting context, cognitive models, and cultures of the workplace. This literature notes that technology must work hand in hand not only with use, but also with context, cognition, culture, training, and management. Two major disciplinary outcomes have been the idea of "co-evolution" [27] and "situated actions" [30]. The first concept tells us that local, contextual practices will shape how a technology is adopted and technology will always change those same practices. Context, cognitive models, and cultures *a priori* will never satisfy these

conditions - there must be some support for a back-and-forth either of prototyping or studying use at various stages.

"Situated actions" highlights two aspects - context and action.

This refigures the user as an actor who makes use of technology in many ways based on situations in which he acts. It implies the autonomy and agency of the user, but only in a rhetorical way. The situations and the actions are still seen as knowledge, which can be obtained in a modestly objective way through a discipline of studying the context of the technology in search for situations and actions rather than uses. Not only that, it subjugates the designer to the technology as well, by marking the agency and autonomy of the designer as immodest. It flips the hierarchy, placing the actor-users as autonomous and above technology, (the technology itself should be as objective as possible), and the designer-user as a modest witness to the user-technology relationship. It sets up an ideal vision of a design process that develops technology for a single-person, single-situation, in which the voice of the designer is muted and the designer is the instrument for obtaining the technology.

Participatory Design makes explicit that the user is a user-subject of the technology-employer alliance. Technology should align itself instead with the user-worker. This acknowledges the knowledge-power of technology itself if it is authored by management. This is a first step in recognizing the question of who should author a system. But it does not acknowledge the power-knowledge of the developer that still exists in Participatory design, i.e. the possession of the technical languages, including *code*.

The first, and most crucial reflection is that the user must be figured into the design process. Even when the designer responds to a call to empower the user or support the user's actions, design of technology always imagines use and figures a "user." Often the rhetoric of Participatory Design and other design methods that are conscious of the conflict, struggles with this tension. Some generalizable use is maintained and this upsets the idealized relationship between the designer and user. Whether the design process refigures the user as a "participant" or an "actor" the word "user" fails to fall away. We would like to consider the "user" as a figure because it enables us to see the design process as a narrative in which the user is the primary figure. An alternative response then is not to remove the user, but to refigure or "mutate" the user but continue to see the user as central to the narrative. We can see how Participatory Design and other user-centered projects are attempts to refigure the user. Often it seems that the new processes have been achieved but that the refiguring of the user has in some senses failed.

"A figure collects up the people; a figure embodies shared meanings in stories that inhabit their audiences" [15]. If we understand current approaches to technology design as narratives we can already see that the user is a figure that collects up the people. The user collects up the selection of people who are seen to be the receivers of technology. While designers may select particular users to involve in the design process, the user-figure clearly points to the collection of all people who will be users, which will inevitably include the

designers themselves. The user-figure embodies shared meanings of use within the contested space of what users and designers make of the technology being designed.

In technology design, the designer has legitimate authority over the object world of designed objects, but this authority does not follow the same "modesty" of the scientific author. The designer, no matter the efforts to empower users, is one who intervenes through the technology s/he creates. The designer is at times expressive through the medium of technological innovation. At other times, especially in much of design methodology, the focus shifts to the user and the designer as a subject recedes from view. However, in what the designer knows and in how s/he intervenes, s/he is immodest. And in the modes through which the designer tries to achieve objectivity (e.g. through a focus on the user), s/he is immodest.

Using the lens of authorship we can further analyze this immodest role of the designer whose voice remains authorial even if there are moves to limit that voice. The immodesty begins to make sense once we realize that design's authorial voice has not yet stabilized in the way that the scientific author has. And designers as a whole are ambivalent, or take conflicting views, on the designer's role as an author, whether to be expressive or "self-invisible." And attempts to place the designer in the self-invisible mode of the modest witness are ultimately unsuccessful because of the technical skill required to author the technology to which, in most cases, only the designer has access.

Literature on Participatory Design and scenarios both consider the need to get software and design professionals to step out of their role and into a new role. This is one way in which the user is clearly a figure, a role that can be taken on by the design professional who chooses to step out of his role. "Can software professionals recognize and affirm the validity of perspectives other than their own, and value the expertise that comes from experience, not just the knowledge that is attested by academic credentials?" [25]. This acknowledges a kind of common sense notion that software professionals need to be disciplined to seek outside perspectives. In *Five Reasons for Scenario-Based Design*, Carroll suggests that scenarios aid in reflection. "Reflection is not always comfortable; it forces one to consider one's own competence, to open oneself to the possibility of being wrong" [4]. This provides a sense that reflection is possible but that reflection of a different kind does not come easily and must be fashioned through discipline and techniques. In Carroll's guide to using scenarios, the designer is figured as a kind of hero - "technical professionals are intelligent people performing complex and open-ended tasks. They want to reflect on their activities, and they routinely do reflect on their activities," but they require an extra push to surmount that higher peak, to reflect on experiences of the non-technical non-professional user. The hero-designer must pander to user experience, user expertise about "experience," and must transcend his role. The user is seen as having expertise about his own experience in both Participatory Design and other

user-centered design. Expertise of experience is a substitute word for subjectivity and still poses the designer as residing in the world of objectivity but needing to reflect on the subjective world in which his work will ultimately reside.

Techniques to step out of the role of designer or to communicate with "real users" are often narrative in form. There are scenarios, which attempt to provide plain English accounts of use in a chronological ordering like a story, or more accurately like one cohesive scene in a story. Scenarios are most often written by developers after meeting with a customer or user and are then provided back to the user or customer for verification or even as a contract for work. "User stories" in Extreme Programming (another software process) are also used as small units of work that are agreed upon in an iterative weekly meeting and which describe features in the language of the user [1].

The discomfort of reflection in the design process is noted and is moderated in the techniques of scenarios and personas, by a mutation of the user into something that the developer has authored and created. The user as a persona or within the plot of a scenario is generic and though this figure can make the designer uncomfortable he is ultimately a phantom of the designer's own mind and thus the designer remains in control. The designer is described as being brought into an "intimate" engagement with the "concrete elements of the situation," the situation that belongs to the user and is obtained by a reflection on the experience of the user. An intimacy with the person who will use the system appears too dangerous to request of the designer. The subjectivity of the user is again removed. The intimacy that the designer should aim to achieve is with "concrete elements of the situation," i.e. something objective about the subjectivity. The designer requires something objective to take away in order to design. This is especially seen in the use of "personas" in the design process. Personas are a literary technique first adopted in marketing and then in product design and ultimately in large-scale software design projects.

Grudin [12] says that personas can be used to make up for some shortcomings of user-centered design and scenarios, particularly in large-scale projects. Often these projects require a specialist group (e.g. academic researchers) to spend time with users, to study, interview and observe them. But Grudin critiques that the data often meant to center design around real users fails to engage the designers.

A solution is personas. "Personas are fictional people. They have names, likenesses, clothes, occupations, families, friends, pets, possessions, and so forth. They have age, gender, ethnicity, educational achievement, and socioeconomic status. They have life stories, goals and tasks" [12]. Designers can author personas without any basis on real users, but Grudin does not recommend this. In many cases personas are authored after extensive data collection about real users as a way to represent the data in a more engaging way. Personas are seen to "take on a life of their own," or "take over" the design process. At the 2007 SIGCHI conference James Nieters of Cisco gave a report entitled "Making Personas Memorable" [26]. Cutout dolls of personas used by

Cisco were provided to conference-goers as take-home souvenirs. Nieters said that one of the benefits of personas is that no one asks for the data that back them up. These very deliberate figurings of the user into "personas" seem to resolve the conflict of broken communication between the user and designer. Personas speak to engineers in a way that real users cannot. But personas are also a way to elide the issue of authorship. The user is kept at a great distance from the design, and is in fact authored into a persona before having a voice in the design process. The designer thus behaves as a ventriloquist for the user as object-world. The persona that the designer authored is inverted in the hierarchy, such that the user-figure is in control of the design but is not a threat to authorship.

Haraway says that the modest witness "is the legitimate and authorized ventriloquist for the object world, adding nothing from his mere opinions, from his biasing embodiment" [15]. The designer fits the role of an immodest witness because he is the "authorized ventriloquist" for the user world, but his opinions and biases are what figure him as the author of the designed technology. This immodesty is a paradoxical dilemma. Unlike the modest witness, the designer is incapable of "self-invisibility" [15]. He must step out of the way for the user to speak the needs of the system, he must ventriloquize this speech modestly, but then he must become visible in the same moment through his authorship of the technology that justifies the design process.

In Participatory Design there is a clear reframing of software design as a political process that involves a conflict of expertise and as a politics of knowledge-creation. It also frames participation in the design process as "not mere 'involvement'" [25]. Muller also acknowledges that participation in the design process is an issue of communication "through language or through artifacts." However, Muller also creates a "Taxonomy of Participatory Design Practice" that poses a dual axis of practice based how early or late the user participates in the design process and how much the user participates in the world of the designer/how much the designer participates in the world of the user.

What is maintained in all these rhetorics is the underlying and under-acknowledged desire of the designer to create "new ways of doing things and new things to do" [4]. In distancing the designer from his desire, design methods create a tension between relinquishing control through in order to understand the real situation of the users and retaining an authorial vision, i.e. the authorship of the design. But this sets up a conflict that cannot be resolved. How can a designer create new ways of doing things and new things to do if they are accountable to the current ways that users do things and the current things that users do.

The role of the designer can be seen as one of a certain expertise in conflict with a different expertise of the user. But there is more to the role of the designer than this kind of expertise. Ultimately designers are good at envisioning use, at imagining and bringing into being new uses and new ways to use. While the framing of design as embedded within

economic and political projects allows the designer to consider the obligation to empower the user as a worker-subject, it does not allow us to consider the conflict of authorship. One can involve the user in the design process at any stage and can migrate the designer to the world of the user or the user to the world of the designer but the authorship of the technology remains an unresolved conflict.

12. IMPLICATIONS: TRANSLATION OR REFIGURATION

This reframing of technology design as Technical Authorship leads to the breakdown of the designer into the roles of expert, author, and especially translator between the technical and non-technical or popular languages. It lays out a different set of obligations and accountabilities for the designer of technology. Not only can we work to make the "user" and the "designer" subjects in design practice, technical experts could succeed in bridging the gap between designers and users by defining a new obligation to translate technical languages and promote technical literacy.

However, Haraway also warns us that "there is no way to rationality - to actually existing worlds - outside stories, not for our species anyway." That is, we should be wary of de-figuring as a route to co-authorship. Within science studies she suggests that we should not unmask the modest witness. Storytelling is the only way that we know about the natural world and so the figure of the modest witness affords us a place to begin telling stories about the world. If we strip the empirical individual of his figures, how will he continue his practice? All the more so, we should be wary of tossing out the immodest witness or user-figure. If storytelling is the only way we can understand the natural world, as Haraway contends, then it is most certainly the only way that we can come to conceive of future worlds we wish to create.

Re-figuring the immodest witness is important in order for reflective and critical design practices to continue to emerge and grow. And the user-figure should certainly not be dealt with rashly because he is design's primary figure that distinguishes design from science. If the sciences had a figure outside of Nature herself to contend with, perhaps the sciences would be better off. That is not to say that the user-figure does not need re-shaping. But perhaps we should embrace the creative methods found in personas and the "user stories" of agile software development as ways to refigure and reshape design.

By reframing the conflict of the user-designer relationship as one of authorship we find a way that the user and the designer can forge a different relationship. This aim aligns with projects for user-centered design and participatory design that seek a non-instrumentalized user-subject but also departs from these through the consideration of the user-figure as central and indispensable to narratives of technical futures. Through the mutation of the user-figure, the qualities that divide the designer-figure from the user-figure fall into several roles that are commonly conflated: that of the author, expert, and translator.

While the politics of the user has been highlighted in various

rhetorics, these have failed to acknowledge the conflict as one of contested authorship. These projects that explicitly express concern for the user, nonetheless maintain authorial control of the technology in the hands of the designer. If the user is the primary figure of the authored narrative, then the user as user-subject stands in the way of authorship, while at the same time being the instrument for discovering an objective "use," and the one whose needs must be satisfied. This is an impossible position for the designer, who is lead into a figurative hero's journey.

By considering the conflict of the user-designer relationship as one of authorship, we are able to highlight the user as the primary figure of the authored work. Current design methodologies can be examined to understand this figure: the "user." This reframing provides the possibility of considering the co-authorship of technology not as the co-design of technology. It leads to the breakdown of the designer into the roles of expert, contested author, and translator between the technical and non-technical languages. It lays out a different set of obligations and accountabilities for the designer of technology.

Co-authorship of technical futures retain the user as the central figure but place the non-designer as a co-author of the narrative the user-figure inhabits including the authorship of a mutated user-figure. The designer plays the role of technical expert, but also translator. The reframing of the problem as one of authorship reminds us that expertise is not the only mode of authorship, that authorship is not the only means for the designer to play a role, and that translation is vital to the political project of building technical futures.

In all cases where the user is figured into the design process or is refigured as an actor, expert, knowledge-maker, meaning-maker, or participant, the attempt to avoid instrumentalization fails to some degree. This is because the user is a figure of technology design, of technology, of the technical and techne. The user is a figure, the focal character in a narrative of imagined future uses. No design process can fail to imagine future uses.

Practices of technology design that do not instrumentalize the user are seen in approaches where authorship is highlighted in the case of artistic or critical technology design where the user is a figure and not intended to inhabit a real person. It is also seen in the abandonment of the user when designers design for designers as is seen in the hackers and opensource movements. The opensource movement, as Kelty [20] discusses in his article on recursive publics, tends towards a mutation technology authorship. Copyleft and other opensource practices are anti-authorial.

Extreme Programming [1], which is an approach to software development, was fashioned as a set of practices to empower software developers in the face of too many demands and requirements from users. It is not often highlighted that this approach also takes a stance against authorship of code. A key principle is that the code belongs to everyone.

Additionally stories are used to translate needs of users into test cases and test cases translate easily into code.

Grudin points out that personas are effective because they are

"a technique that... can draw upon powerful psychological forces to restore these dimensions" of engagement, complexity, representation, and identity. In this, and his discussion of fiction, Grudin acknowledges what Haraway says that "there is no way to rationality-to actually existing worlds-outside stories, not for our species, anyway" [11]. But the success of personas and other kinds of narrative devices in the design process should not be used as a wedge between the user and the developer. Instead this should form a recognition that the user-figure is central and is always authored with or without data collection.

The user figure should not be put to sleep but should be "mutated" as Haraway suggests with the mutation of the sciences' modest witness. This mutation happens through the focus on technology design as technology authorship. Adrian Johns discusses how authorship in the sciences was crafted over time. The "modern authorial persona" was shaped through "shifts in policing, property, bibliographical classification, and, finally... understandings of the creative process" [19]. In design, the creative process is one of the primary narratives, and so it is not surprising that fictional methods have been deployed in design while they have been extricated from the scientific authorial voice.

Stories are already a part of design, but they are "fictional methods to convey profound truths" [19], when they ought to be fictional methods to convey profound fictions. Those who write about a critical technical practice in technology design such as Phoebe Sengers, draw attention these narratives of future use. But the user-figure is not central to these narratives as much as the technology which instantiates future uses that were previously unknown.

The user-figure is mutated from its original role because it is now distinct from the empirical users, is maintained as the central figure, is more radically subjective, but also intersubjective. The cat's cradle that Haraway discusses requires a back and forth between the designer and non-designer in shaping knots and paths that will make up the narrative of technical futures [15]. The co-authors are the players of the game and should not be tied up in its knots. The user figure is the use-maker, to draw on the idea of knowledge-maker, he is the character in a story of rain-making, the one who brings about rain by inhabiting a story that imitates the rain and is the rain. The user-figure can be to the modest witness what the white rabbit is to Alice, or the fool to the magus, the one who is followed through the story to tell the story itself.

13. REFERENCES

- [23] Beck, K. 2000. *Extreme Programming Explained: Embrace Change*. Addison-Wesley Professional.
- [24] Beck, K. et al. 2001. "The Agile Manifesto" <http://agilemanifesto.org/>
- [25] Biagioli, M. and Galison, P. 2003. "Introduction." In *Scientific Authorship: Credit and Intellectual Property in Science*, M. Biagioli and P. Galison, Eds. Routledge, New York and London, 1-9.

- [26] Carroll, J.M. 2000. Five reasons for scenario-based design. *Interacting with Computers* 13, 43-60.
- [27] Chartier, R. 2003. Foucault's Chiasmus: Authorship between Science and Literature in the Seventeenth and Eighteenth Centuries. In *Scientific Authorship: Credit and Intellectual Property in Science*, M. Biagioli and P. Galison, Eds. Routledge, New York and London, (Chap 1) 13-31.
- [28] Coney, M.B. and Chatfield, C.S. 1996. Rethinking the author-reader relationship in computer documentation. *ACM SIGDOC Asterisk Journal of Computer Documentation* 20, 23-29
- [29] Elliott, M. and Scacchi, W. Mobilization of Software Developers: The Free Software Movement. *Information, Technology and People*.
- [30] Erickson, T. 1996. Design as storytelling. *interactions* 3, 4 (Jul. 1996), 30-35. DOI= [doi.acm.org—234813.234817](https://doi.org/10.1145/234813.234817)
- [31] Foucault, M. 1977. "What is an Author?" In *Language, Counter-Memory, Practice*. D. F. Bouchard and S. Simon, Trans. Ithaca, New York: Cornell University Press, 1977. pp.124-127.
- [32] Gallardo-Valencia, R.E. and Sim, S.E. 2007. "Planning and Improvisation in Software Processes," *ACM Crossroads*, Volume 14. Issue 1.
- [33] Grudin, J. 2003. The West Wing: Fiction Can Serve Politics. *Scandinavian Journal of Information Systems* 15, 73-77.
- [34] Grudin, J. and Pruitt, J. 2002. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. *Proc.PDC* 144-161.
- [35] Grudin, J. 1993. Interface: an evolving concept. *Commun. ACM* 36, 4 (Apr. 1993), 110-119. DOI= [doi.acm.org—255950.153585](https://doi.org/10.1145/255950.153585)
- [36] Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J. and Cajander, Å. 2003. Key principles for user-centred systems design. *Behaviour & Information Technology* 22, 397-409.
- [37] Haraway, D.J. 1997. *Modest-Witness@ second-Millennium. Femaleman-Meets-Oncomouse: Feminism and Technoscience*. Routledge Chapman & Hall.
- [38] Holtzblatt, K. 2002. Personas and contextual design. Retrieved August 17, 2003.
- [39] Jackson, M. 2003. Can Artisans Be Scientific Authors? The Unique Case of Fraunhofer's Artisanal Optics and the German Republic of Letters. In *Scientific Authorship: Credit and Intellectual Property in Science*, M. Biagioli and P. Galison, Eds. Routledge, New York and London, (Chap 5) 113-131.
- [40] Johansson, M. and Messeter, J. 2005. Present-ing the user: constructing the persona. *Digital Creativity* 16, 231-243.
- [41] Johns, A. 2002. The Ambivalence of Authorship in Early Modern Natural Philosophy. In *Scientific Authorship: Credit and Intellectual Property in Science*, M. Biagioli and P. Galison, Eds. Routledge, UK, (Chap 3) 67-90.
- [42] Kely, C. 2005. Geeks, Social Imaginaries, and Recursive Publics. *Cultural Anthropology* 20, 185-214.
- [43] Kyng, M. 1998. Users and computers: a contextual approach to design of computer artifacts. *Scandinavian Journal of Information Systems* 10, 7-43.
- [44] Latour, B. 1994. *Science in action: how to follow scientists and engineers through society*. Harvard University Press, Cambridge, MA.
- [45] Lessig, L. 1999. *Code and Other Laws of Cyberspace (Code v2 Wiki)*. Basic Books.
- [46] Madsen, K. H. and Aiken, P. H. 1993. Experiences using cooperative interactive storyboard prototyping. *Commun. ACM* 36, 6 (Jun. 1993), 57-64. DOI= [doi.acm.org—153571.163268](https://doi.org/10.1145/153571.163268)
- [47] Muller, M. J. and Kuhn, S. 1993. Participatory design. *Commun. ACM* 36, 6 (Jun. 1993), 24-28. DOI= [doi.acm.org—153571.255960](https://doi.org/10.1145/153571.255960)
- [48] Nieters, J. E., Ivaturi, S., and Ahmed, I. 2007. Making personas memorable. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems (San Jose, CA, USA, April 28 - May 03, 2007)*. CHI '07. ACM, New York, NY, 1817-1824. DOI= [http://doi.acm.org/10.1145/1240866.1240905](https://doi.org/10.1145/1240866.1240905)
- [49] Rogers, Y. 1994. Exploring obstacles: integrating CSCW in evolving organisations. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (Chapel Hill, North Carolina, United States, October 22 - 26, 1994)*. CSCW '94. ACM Press, New York, NY, 67-77. DOI= [http://doi.acm.org/10.1145/192844.192875](https://doi.org/10.1145/192844.192875)
- [50] Rheinberger, H. 2003. "Discourses of Circumstance" A Note on the Author in Science. In *Scientific Authorship: Credit and Intellectual Property in Science*, M. Biagioli and P. Galison, Eds. Routledge, New York and London, (Chap 12) 310-323.
- [51] Shapin, S. 1994. *A Social History of Truth: Civility and Science in Seventeenth-century England*. University of Chicago Press, Chicago.
- [52] Suchman, L.A. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press,
- [53] Vredenburg, K., Mao, J., Smith, P. W., and Carey, T. 2002. A survey of user-centered design practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves (Minneapolis, Minnesota, USA, April 20 - 25, 2002)*. CHI '02. ACM Press, New York, NY, 471-478. DOI= [doi.acm.org—503376.503460](https://doi.org/10.1145/503376.503460)

FULL PAPER SUBMISSION 5:

The d-school in the i-school: HCI and Design Research

AUTHOR(S):

Carroll, J. M.
 Haynes, S. R.
 Ritter, F. E.
 Rosson, M. B.
 Zhang, X. L.

ORGANIZATION(S):

Pennsylvania State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

In the context of a campus-wide strategic planning process, we have been asked to frame the special role of Penn State's Center for Human-Computer Interaction (CHCI) within the College of Information Sciences and Technology (IST). Our response is that HCI should fulfill an important role as a focus for *design research* within our i-school. As support for this response we explore the following propositions:

1. Design research and design science are essential to i-schools.
2. HCI is historically, currently, and inherently a design science of the just right sort to contribute to i-schools.
3. Other constituencies/subunits of i-schools (CS, IS, LIS, STS) are not design sciences, but benefit from organizational proximity of HCI design research.

In this brief paper, we sketch the argument entailed by these propositions. Our objective is to better understand and articulate the role of HCI and design research in i-schools.

Design research is essential to i-schools

I-schools span the information disciplines to comprehensively address issues ranging from image processing to social policy. In this vast spectrum, design - defined here as *materializing information to support human activity* - is a fulcrum. One cannot talk about information processing, storage and retrieval, information behavior, the use of information in groups, organizations, or in society, or information economics, policy, and regulation without making strong and substantive assumptions about the specific ways that information can be presented to and manipulated by people. Indeed, whenever there is evolution in user interfaces and application services, other concepts and relations throughout information science must be rethought.

There are perfectly coherent and time-tested information disciplines that eschew a central concern with design as we define it. The core of Computer Science (CS) is focused on information processing, storage and retrieval; presentation and manipulation of information by people, and the design process itself, are peripheral topics in CS. Library Information Science (LIS) has traditionally focused on information retrieval and information seeking at the individual level, while Information Systems (IS) has focused on organizational processes and the impacts of information access and use. But neither has focused on framing and implementing specific design strategies to ameliorate or enhance observed circumstances through transforming human-computer interactions.

The broad vision of i-schools requires coordination of understanding information structures, needs, and impacts with the design of new human-computer interactions. Design and the artifacts that result from design activity represent an important class of boundary objects for knowledge and technology transfer between disciplines and between academe and industry/government. Designs integrate and reify theory and empirical knowledge into technologies that can be exchanged, implemented, and their effects assessed within different contexts and from different perspectives. Similarly, designs and artifacts help demonstrate the potential utility of research products.

HCI is the right kind of design science

Until the emergence of HCI during the 1980s, the roles for social, cognitive and behavioral science within technology development were few and marginal. The discipline of human factors, institutionalized within Industrial Engineering or Applied Psychology, emerged a generation before HCI, but it was most concerned with human performance and experience with existing technology tools and systems, and focused on artifacts with limited information content (e.g., stairs, furniture, physical tools). This *evaluation* role was often positioned too late in the system development process to have more than minor effects on the design of the technology or on the core applications of technology.

Science and Technology Studies (STS), institutionalized variously in philosophy and sociology; LIS, typically a standalone academic unit, some of which have developed into i-schools; Communication Studies, and IS, institutionalized in business schools; all have provided descriptions of how technology is utilized across a broad range of significant organizational and utilitarian contexts. Particularly in the case of STS, these were often rich analyses of how new technology is shaped by use, and how it reciprocally shapes human practices, relationships, and institutions. Such efforts contribute much to understanding the context of information technology design problems and solutions, but they are even more distant from the activities of system design and development than human factors efforts. And unfortunately technologists were largely unaware that these fields of interest even existed.

In this context, HCI is quite novel. It has been most often institutionalized as an adjunct to, sometimes even within, Computer Science. Its research methods integrate the creation of new techniques, technology, and applications with iterative and formative evaluation of their immediate and indirect impacts on collective and individual human activity and experience. The close coordination of creating innovative technology and assessing its consequences for human activity has proven very successful, and has led to transformative applications—desktop user interfaces, the mouse and carefully engineered pointing devices, hypermedia online information, a variety of information visualizations, high quality synthetic speech, and levels of user control of software systems that were unimaginable two decades ago.

Indeed, a paradigm of innovation and application of interactive technologies cyclically coupled to improvisation and assessment in human activity has been incorporated into academic programs in geo-science, IS, CS, LIS, and i-schools, and is generally called HCI. However, this paradigm is not itself a discipline, and when it is incorporated into specific information technology disciplines and projects it is typically *narrowed to the scope of that discipline or project*. For example, HCI in CS most typically means research on software architectures or tools that enable new user interface techniques. We propose that *only in i-schools* can HCI operate with the appropriate scope to be effective as a design science (see also Carroll et al., 2006).

Other i-school constituencies benefit

Incorporating HCI into i-schools yields many benefits for other i-school constituencies, while at the same time broadening and enriching HCI design research. Already the combination of HCI and LIS in i-schools has led to richer conceptions of human-centered design that encompass both the measurement orientation of HCI and the integration of information, organization, technology, policy, and culture of LIS. The interaction of HCI with community informatics has promoted design research that addresses new sorts of public access information systems, and has produced a more finely articulated view of people as the central factor in the interactions between people, information, and technology at different levels of social aggregation.

Traditional CS research considers underlying technologies but rarely addresses design in the sense of envisioning new tools and activities for people. Instead CS design efforts are more likely to lead to self-consistent functionality with unknown implications for people and their activity (e.g., a programming language), or perhaps to a proof-of-concept prototype application that demonstrates a plausible use of a novel technology or technique. Incorporating HCI expertise in such work ensures that the design of effective applications is a first-order concern, and helps to coordinate the development of new technologies with new measurement techniques they necessitate.

IS has an inherent orientation toward action science and affecting change, but generally has accomplished this through organizational interventions of training, management, and technology acquisition. The ability to design and implement new forms of human-computer interactions as part of IS interventions is a powerful enhancement. Other constituencies, like STS, that have traditionally played even more peripheral roles with respect to technology (e.g., focusing on impacts or critically deconstructing meanings) can participate in a more directive fashion when supported by colleagues who do this as a central element of their technical work.

Our construal of the role of HCI in i-schools - as an anchor for design research - has more general implications for how we think of i-school faculty groups. For example, at Penn State, we have often identified faculty groupings concerned with "people and technology", "people and information", or "information and technology". Not surprisingly, such distinctions do little to energize engagement across groupings. Perhaps differentiating faculty groups based on *how* they pursue research questions in the i-space (which is a nexus of information, technology and people), and equally important, *how* their methods and approaches can complement others would be more productive.

Reference

Carroll, J.M. Dourish, P., Friedman, B., Kurosu, M., Olson, G.M. & Sutcliffe, A.G. 2006. Institutionalizing HCI: What do i-schools offer? Extended Abstracts of ACM CHI 2006: Conference on Human Factors of Computing Systems (April 22-27, Montreal). New York: ACM Press, pgs 17-20.

FULL PAPER SUBMISSION 6:

Tracing Transnational Flows of IT Knowledge Through Open Exchange of Software Development Know-How

AUTHOR(S):

Sim, Susan Elliott

Philip, Kavita

ORGANIZATION(S):

University of California, Irvine, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Introduction

Technology is often promoted as a means to rapidly advance developing nations. In the current era, information technology (IT) is commonly touted as the salvation of developing economies. India is one country that has embraced this notion. In the pages of American business magazines, pathos-filled stories about underdeveloped subjects and the objects of aid policy give way to, or at least share space with, excited narratives of hyper-developed subjects with amazing computation skills. An activist-programmer we recently interviewed in Bangalore concurred with this narrative, stating "IT is the only thing that can move India from a resource-poor to a knowledge-rich economy."

Despite this promise, the challenge of sharing technology through transnational circuits is non-trivial. Technological exchanges and translations must span geography, cultures, languages, configurations of resources, and local contingencies. The question then arises, what are appropriate vehicles for sharing the IT knowledge needed to develop innovative and effective software? Is Open Source the great disseminator of intellect? Are social connections integral to this, and if so, what forms? We tackled these questions in an interdisciplinary research project combining history of technology, contemporary ethnography, and software engineering. We conducted fieldwork in Bangalore, India to understand the sharing practices that take place in transnational flows of software development knowledge. Working from historical and ethnographic studies, we outline models and predictions regarding optimal modes of collaborative transnational software development.

In this paper, we argue that Open Exchanges of software development know-how are necessary to create a cadre of information workers who are capable of developing future innovative and practical software applications. By incorporating software developers who are on the periphery, the global community of practice is broadened, which results in mixed modernities that are capable of meeting the needs of developing nations.

This paper is part of a group three related papers being submitted from the Department of Informatics at University of California, Irvine. These papers address the theme of bringing the promise of IT to under-addressed groups. This paper examines the participation of software developers from India in a global community of practice. A second paper considers the problem of bringing mobile computing to infrastructure-poor regions, such as South Africa. The third and final paper uses the metaphor of technical authorship to frame the power and knowledge mismatches between non-technical and technical participants in technology design. In addition to their thematic commonality, these papers takes a broader view than is typical in software engineering and human-computer interaction, and includes societal and historical considerations.

Role of Know-How in Software Development

The acquisition of know-how is necessary to attain proficiency in the craft of software development. Whereas know-what is explicit, factual knowledge, know-how is the ability to put know-what into practice. Know-how is necessary to successfully create innovative products. Programming languages, application frameworks, and software tools are general-purpose technologies. They are intentionally designed to be highly flexible and adaptable, and the onus is on the developer to use these tools to create specific solutions. Furthermore, this know-how needs to be constantly updated, because software technology and the information ecologies of end-users are constantly changing.

Open Source software is often cited as means for leveling inequalities between intellectual and material haves and have-nots. Unfortunately, it is very difficult to share know-how through artifacts due to the properties of source code. Program source is too complex to be understood on its own (or even with typical documentation), too brittle to travel well between settings, and contains completed solutions, but not the know-how needed to build new solutions. Many technologies are complex, but software especially so, and this complexity is often compounded by its size. It can be very difficult to locate the know-how in this tangled mass of classes, interfaces, design patterns, delocalized plans, and scattered concerns spread across millions of lines of source code. As well, it can be difficult to modify, since the omission of a single file, or even an error in a single character can cause a program to fail to compile into an executable.

Open Source is necessary, but not sufficient to transmit know-how. This knowledge must be acquired through experience or from other practitioners.

Open Exchanges for Know-How

An open exchange is a space where interested people can learn, critique, and contest ideas. It is not so much a physical or virtual place, or even a particular event, but rather an opening for a particular kind of interaction. It's a locale that is defined in terms of activity, rather than geography, technology, or membership. Another formulation of an open exchange in terms of activities is "link, lurk and try," meaning linking with others of like minds, lurking on the periphery of a community of practice, and trying out new things with low risk (Brown and Duguid 2002). Examples of "trying" are accessing and using new technologies, airing new ideas, and rehearsing arguments.

Any site that has the capability for a community of individuals to interact with each other directly has the potential to become an open exchange. What distinguishes a merely social group from an exchange is common practice, that is, the organization of the community around a set of problems, technologies, or know-how. John Seely Brown has argued that trust and "creative abrasion" in such communities are key (Brown and Duguid 2002).

The Open Source movement has resulted in a worldwide community of practice and a network of open exchanges (Goldman and Gabriel 2005; Lave and Wenger 1991). Communities of practice are a highly effective means for learning know-how, especially in domains involving design and technology (Fischer 2004). The project source code, discussions in electronic forums, and solutions in the form of bug tracking and change sets is open for anyone to examine, comment on, and contribute to. Transmission of know-how occurs through both active participation and legitimate peripheral participation, or lurking.

BarCamp Bangalore

One of our illustrative examples is the mode of know-how exchange that happens at BarCamps in Bangalore. Around the world, BarCamps are ad hoc gatherings of software developers, explicitly formulated as an open, interactive exchange. Forged in opposition to perceived exclusions in the sharing of software development know-how, BarCamps, often referred to as "non-conferences," challenge the hierarchies among speakers and audience, experts and laypersons. Organized in the form of spontaneous collectives, BarCamps exhibit *ad hoc* community exchange that employs the most flexible current tools of know-how exchange, including wikis, wifi, social bookmarking, photosharing, blogging. They have become highly popular "semi-official" channels where software developers learn from, and forge, communities of practice.

BarCamps all over India have been vital in projects such as localization of software, popularizing new programming languages and techniques among non-native English speakers, and the discussion of the social and political context of emerging IT economies. One of us (Philip) attended BarCamp Bangalore (BCB) (BarcampBangalore 2007) in August 2007 to learn about open exchanges in this context.

BarCamps are full of a palpable excitement. Participants give up weekend leisure (after grueling work-week schedules in corporate programming jobs) in order to meet with people they come to consider their most intimate community, yet whom they largely know only on-line. The combination of virtual and physical worlds, technical and social discussions, work and leisure, and multiple programming and human languages makes BarCamps thrilling examples of open exchange for participants. One of the main organizers expressed his commitment to open public exchanges of knowledge as the primary reason for his devotion to the BarCamp project.

Although BCB participants are not identical to the Open Source community, many of them spoke of their involvement in India's free software movement. Many BCB participants reported searching for open exchange forums not primarily from ideological opposition to proprietary software, or social commitments to transparency, but simply because they found they were not learning rapidly enough in closed systems.

The emergence of open exchanges such as BCB serve to underline our critique that the transmission of Open Source software artifacts is not sufficient to transmit know-how. As well, social networking in isolation is not sufficient to create the communities of practice that are necessary to locate and share know-how. Rather, it is the configuration of these elements into open exchanges that creates a cadre of software developers that is needed to develop the software needed by local contexts. In this manner, IT knowledge flows across national boundaries to bring information workers and countries into a global community.

References

- Brown, John Seely, and Paul Duguid. 2002. *The Social Life of Information*: Harvard Business School Press.
- Fischer, Gerhard. 2004. *Social Creativity: Turning Barriers into Opportunities for Collaborative Design*. Paper read at Eighth Conference on Participatory Design, 27-31 July 2004, at Toronto, ON, Canada.
- Goldman, Ron, and Richard P. Gabriel. 2005. *Innovation Happens Elsewhere: Open Source as a Business Strategy*: Morgan Kaufmann.
- Lave, Jean, and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*: Cambridge University Press.
- Main Page - BarcampBangalore 2007 [cited 21 September 2007. Available from http://barcampbangalore.org/wiki/Main_Page.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR

Tracing Transnational Flows of IT Knowledge Through Open Exchange of Software Development Know-How

Susan Elliott Sim
Department of Informatics
University of California, Irvine
Irvine, CA 92697-3440
+1 949 824 2373
ses@ics.uci.edu

Kavita Philip
Department of Women's Studies
University of California, Irvine
Irvine, CA 92697-2655
+1 949 824 7092
kphilip@uci.edu

ABSTRACT

Information technology (IT) is often promoted as a socially and culturally agnostic tool that will allow emerging economies to leap into the digital age and reap the wealth that accompanies it. But in addition to the programming language, software tools, and books, *know-how* is needed to turn bright ideas into innovative, marketable solutions. This know-how can only be acquired from experience or from other IT developers. An effective means for sharing know-how is through an open exchange, which we characterize as a space where interested people can learn, critique, and contest ideas. It's a locale that is defined in terms of activity, rather than geography, technology, or membership. We use open exchanges of know-how as an analytical lens to example historical examples and contemporary instances. While conducting fieldwork in India, we observed open exchange occurring at Barcamp Bangalore. It is through these exchanges of know-how, rather than the transmission of tools or software artifacts that IT knowledge flows between international locales. We conclude this paper with a discussion that is mutually informed by contemporary practice and historical configurations.

Topics

Community technologies and networking
Information technology and services

Keywords

Open Source, software development, knowledge, know-how, transnational circuit

1. INTRODUCTION

Information technology (IT) is commonly touted as the route to national progress and the renewal of the global economy. While there are many good reasons to accept this claim at face value, we suggest that future IT innovation might better be served by

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse this work. The authors are solely responsible for their paper's content. Our thanks to the Association for Computing Machinery for permission to adapt and use their template for the iSchools 2008 Conference.

analyzing this process (a) technologically and (b) socio-historically. A more nuanced formulation of this claim is likely to have beneficial effects on various kinds of policy-making that base themselves on versions of this claim.

First, we wish to disaggregate the notion of knowledge itself, and specifically IT knowledge, into two components, "know-what" and "know-how." No doubt IT could be further subdivided, but at a first level of approximation, this heuristic division reminds us that technology itself is both a knowledge-system and a practice. It is both a highly systematized, explicitly formulated and repeatedly tested set of logically nested truths (which allows us to *know what* laws hold), and a highly fluid set of practices, intuitive beliefs, and implicit codes (which allow us to *know how* to do things).

There has been much historical and sociological work on expertise (and "know-what" forms of IT knowledge), but less on the intuitive, socially coded set of practices we wish to lump under the heading of *know-how*. We suggest that understanding know-how will help us better understand some of the ways in which (1) groups of software developers share knowledge, (2) software development innovates, and, (3) IT knowledge travels along transnational circuits of practitioners.

Thus the policy areas that are elucidated include questions such as: Where should we look for the next wave of ideas in software development to fuel the nation's economy? We suggest in closing that many sites of often-overlooked creativity and innovation may exist in marginal networks outside the mainstream areas of software development.

Below, we briefly explain what we mean by know-how. We then explore why a more complex ethnographic understanding of know-how might elucidate the ways in which software innovation occurs, and how this shifts our models of global dispersion of technological practices. Finally, we explore the policy implications.

2. WHAT IS TECHNOLOGICAL KNOW-HOW?

In their classic work "The Social Shaping of Technology," Donald Mackenzie and Judy Wajcman [1] suggest that technology has three layers of meaning: physical objects, the human activities associated with these objects, and most importantly, knowledge about how to conceive, design, build and repair these objects, a rather fuzzy area they term know-how. This latter area, "know-

BELOW)

how," is fuzzy, and not merely to *post hoc* analysis. More intriguingly, expert technological practitioners often cannot put into words how they know what to do when confronted with a technological challenge. Other theorists have suggested that know-how cannot be captured in words, and is often visual and tactile, not just mathematical or verbal [1] (pp. 3-4), [2]. Historians of technology have suggested that this sort of implicit, practical skill at the nexus of art and craft is in fact the older meaning of technology itself, and have explored the variations of meaning in *la technologie* (French, "technology"), *la technique* (French, "technique"), and *die Technik* (German, "technics," in Lewis Mumford's translation).

A set of issues is raised for us by this STS approach to technology and the practical arts. First, how does know-how function in software development? Second, what can we learn from other historical examples? And finally, how does this function in a global, not just national, context?

2.1 The Role of Know-How in Software Development

Know-how has an important role in software development and its acquisition is necessary to attain proficiency in the craft. Whereas know-what is explicit, factual knowledge, know-how is the ability to put know-what into practice. Know-how is necessary to successfully innovate. Programming languages, application frameworks, and software tools are general-purpose technologies. They are intentionally designed to be highly flexible and adaptable, and the onus is on the developer to use these tools to create specific solutions. Furthermore, this know-how needs to be constantly updated, because software technology and the information ecologies of end-users are constantly changing.

Know-how is the link between creativity and innovation. Creativity can be described as the generation of ideas, while innovation is the practical application of those ideas into workable solutions. Know-how is the procedural and experiential knowledge that is needed to perform the transformation successfully. In information technology, know-how takes many forms, such as working knowledge of application program interfaces (APIs) and libraries, the craft skills for creating a database schema, and the judgment to know when to apply different principles. This IT know-how is highly situated, which means that it needs to be adapted to specific situations, and constantly emerging, because the problems and technologies are constantly changing. Therefore, open exchanges where people can seek out, provide, and share IT know-how are particularly important in this domain for sustaining innovation.

2.2 Open Source Software

The most commonly cited space of open exchange, is the domain of open source developers, who work through communities based on exchanging and increasing know-how. Open Source software is often cited as means for leveling inequalities between intellectual and material haves and have-nots. Although Open Source developers have forged ingenious modes of sharing knowledge, it is in fact surprisingly difficult to share know-how through software artifacts.

This difficulty is due to properties of source code. Program source is too complex to be understood on its own (or even with typical documentation), too brittle to travel well between settings,

and contains completed solutions, but not the know-how needed to build new solutions. The constraints of the problem, the problem context, and the decisions made as part of the design process are missing. Many technologies are complex, but software especially so, and this complexity is often compounded by its size. An application like Microsoft Excel can contain upwards of ten million lines of source code. It can be very difficult to locate the know-how in this tangled mass of classes, interfaces, design patterns, delocalized plans, and scattered concerns. As well, it can be difficult to modify, since the omission of a single file, or even an error in a single line of code can cause a program to fail to compile into an executable that will run correctly.

Open Source software artifacts are necessary, but not sufficient to transmit know-how. Software development know-how needs to be acquired through experience or from other practitioners. Every successful Open Source project has a corresponding community that interacts through forums or mailing lists, and sometimes through embodied meetings. We have found that open exchanges are critical to the exchange and transmission of software development know-how.

3. OPEN EXCHANGES FOR SOFTWARE DEVELOPMENT KNOW-HOW

An open exchange is a space where interested people can learn, critique, and contest ideas. It is not so much a physical or virtual place, or even a particular event, but rather an opening for a particular kind of interaction. In this sense, the classical Greek agora, as a marketplace of ideas, a crossroads for intersecting contingencies, and a forum for a critical public community, was also an open exchange. It's a locale that is defined in terms of activity, rather than geography, technology, or membership. Another formulation of an open exchange in terms of activities is "link, lurk and try," meaning linking with others of like minds, lurking on the periphery of a community of practice, and trying out new things with low risk [3]. Examples of "trying" are accessing and using new technologies, airing new ideas, and rehearsing arguments.

Any site that has the capability for a community of individuals to interact with each other directly has the potential to become an open exchange. They are not necessarily geographic locations or regular events; they can be virtual spaces and informal meetings. Their memberships can be restricted and controlled, or they can be informal and unregulated. The exchanges can be public, or they can be within an organization. What distinguishes a merely social group from an exchange is common practice, that is, the organization of the community around a set of problems, technologies, or know-how. The adjective "open" applies to the sharing of ideas and solutions, unfettered by hierarchical structural constraints, reporting relationships, and professional rank. The most effective open exchanges appear to be ones with a diverse membership, with people representing a broad spectrum of local contingencies and social groups. John Seely Brown has argued that it is the trust and "creative abrasion" in such communities that is the key to innovation [4]. Innovation, he argues convincingly, cannot be "managed," but instead must be critically nurtured by creating a space for pluralism—neither stifling it, nor letting it run amuck.

The Open Source movement has resulted in a worldwide community of practice and a network of open exchanges [5, 6]. Communities of practice are a highly effective means for learning know-how, especially in domains involving design and

technology [7]. The project source code, discussions in electronic forums, and solutions in the form of bug tracking and change sets is open for anyone to examine, comment on, and contribute to. Transmission of know-how occurs through both active participation and legitimate peripheral participation, or lurking.

3.1 Historical Examples

There are many historical and contemporary examples of innovation that has been fueled by open exchanges.

Libraries are early examples of sharing. The great libraries of Alexandria, Tunis, Tibet, Nalanda, Baghdad were centers of shared learning to which scholars would travel, sometimes for years, to avail of free knowledge. A revolution in cataloguing during the 1800s revealed a dense web of connections among multiple knowledge elements and processes, "transforming the library catalog from an inventory into an instrument of discovery" [8]. By identifying these connections and innovative combinations of knowledge elements, the library became an open exchange and fostered knowledge creation.

Open exchanges have followed global transnational circuits long before the modern era. For centuries, preachers, traders, warriors and adventurers carried shared experiences, ideas, and memories around the globe, creating global markets and shaping networked histories [9]. While some analyses have characterized these flows as traveling between cores and periphery, the story is a more complicated one than the standard one about conquest, domination, submission, and tribute. Rather, transnational circuits of commerce and culture became the conduits for the later development of sharing, hybrid networks. Today, open exchanges for know-how within transnational circuits are central to innovation in software development.

4. Open Exchanges of Software Know-How in India Today

Around the world, BarCamps are *ad hoc* gatherings of software developers, explicitly formulated as an open, interactive exchange [10]. Forged in opposition to perceived exclusions in the sharing of software development know-how, BarCamps, often referred to as "non-conferences," challenge the hierarchies among speakers and audience, keynotes and panels, experts and laypersons. Organized in the form of collectives that meet on the fly, BarCamps exhibit a mode of *ad hoc* community exchange that employs the most flexible current tools of know-how exchange, including wikis, wifi, social bookmarking, photosharing, blogging, and chat. They have become one of the most popular "semi-official" ways in which software developers learn from, and forge, communities of practice.

BarCamps all over India have been vital in projects such as localization of software, popularizing new programming languages and techniques among non-native English speakers, and the discussion of the social and political context of emerging IT economies. One of us (Philip) attended BarCamp Bangalore (BCB) [11] in August 2007 to learn about open exchanges in this context.

BarCamps are full of a palpable excitement. Participants give up weekend leisure (after grueling work-week schedules in corporate programming jobs) in order to meet with people they come to consider their most intimate community, yet whom they largely know only on-line. The combination of virtual and physical worlds, technical and social discussions, work and leisure, and

multiple programming and human languages makes BarCamps thrilling examples of open exchange for participants. One of the main organizers of BCB expressed his commitment to open public exchanges of knowledge as the primary reason for his devotion to the BarCamp project.

Although BCB participants are not identical to the Open Source community, many of them spoke of their involvement in India's free software movement. Many BCB participants reported searching for open exchange forums not primarily from ideological opposition to proprietary software, or social commitments to transparency, but simply because they found they were not learning rapidly enough in closed systems.

The emergence of open exchanges such as BCB serve to underline our critique that the transmission of software artifacts through the Open Source movement is not sufficient to transmit know-how. As well, social networking practices and technologies are not sufficient to create the communities of practice that are necessary to locate and share know-how. Rather, it is the configuration of the elements into open exchanges that is necessary to create an autonomous cadre of software developers that is needed to develop the software needed by local contexts. In this manner, IT knowledge flows across national boundaries to bring information works and countries into a global context.

5. Historical Lessons, Revisited

Is today's Open Source revolution another manifestation of the parameters that have historically governed technological progress?

Our preliminary investigations suggest that, although the current period of IT innovation shows many novel aspects, there are strong historical resemblances, which are worth investigating further. For example, historian Pamela Long's study of "Technical Arts and the Culture of Knowledge from Antiquity to the Renaissance" [12] finds that in the fifteenth- and sixteenth-century, technological practitioners commonly advocated openness, freely disseminating their knowledge of subjects such as mining, ore processing, artillery, and fortification. Other historians of science have shown that early modern scientific communities were rooted in "moral economies," formed by webs of values about collaboration, self-discipline, and sharing [1, 13]. These values were initially drawn from the ambient culture, such as the model of the humble, dedicated, and self-disciplined saint, which influenced seventeenth century models of the natural philosopher, and remained intrinsic to eighteenth century scientific personality-ideals. These values, however, morphed over time into specific scientific values without direct correspondences to the broader culture.

Scientific cultures reworked and re-circulated a dynamic web of values to form moral economies particular to each scientific community. The resulting scientific "moral economy" could not be explained simply as a reflection of cultural norms. More recent ethnographic studies of "hacker" communities suggests that their technological practice is closely tied to their "cultural" values regarding freedom, individuality, sharing, and innovation. For example, ethnographer Gabriella Coleman suggests that the form and content of Open Source software embodies structures of linkage, transparency and connectivity that are dynamically related with similar values in the Open Source community [14]. Since Open Source communities are globally dispersed, and function without conventional face-to-face interaction, there are few social mechanisms to enforce structures of exchange. Rather

than cultural or organizational conventions, it is technology that functions as the medium of exchange, embodying "values" of openness.

The mutual embeddedness of open cultures of exchange and technologies of open design is only just beginning to be understood. We suggest that the combination of historical, ethnographic, and software engineering methodologies provides a promising route to a robust understanding of this important dynamic, which is a part of the most innovative new software design practices today.

Techniques and practices of sharing are some of the most creative, and the most controversial, technological developments of the last decade. It is the shared excitement about technological challenges that facilitates, and motivates, new kinds of remote communication. It is the speed and global scope of the Internet that allows young "geeks" to build global shared communities, and any perceived threat to widespread access fuels much of their activism [15-18].

Why do software sharing communities, or what we call know-how agoras, spend so much time and energy developing technologies and practices of sharing? Many Open Source developers have day-jobs in proprietary software companies, but spend their personal, unpaid hours coding and de-bugging software that belongs to nobody, yet is potentially anybody's. A shared excitement about dispersed problem solving, as well as a commitment to open exchange of IT knowledge, fuels Open Source communities. Compartmentalization (into departments and companies) appears to be wasteful (in terms of optimizing creativity), because knowledge is "siloed," meaning there is a narrow division of labor. Developers spend a lot (too much) of their time locating resources, and not enough time playing with the resources. Theorists across disciplines, including legal anthropologist Rosemary Coombe [19], ethnographer of globalization Arjun Appadurai [20], and scholar-entrepreneur John Seely Brown [4] have pointed out that information has a cultural and social life—that is, it is produced, shaped, exchanged, and designed in fundamentally social contexts, not just in individual minds. Sharing is a key necessity for this transformation in the patterns of design innovation, and the emergence of open exchange systems for software know-how is a key to creativity in IT.

6. Discussion and Implications

Histories of science and technology teach us how to look for webs of values particular to extremely innovative technical communities. These histories also indicate why an interdisciplinary method is necessary: while historians of culture and ethnographers might be alert to forms of communication, patterns of collaboration, and innovative organizational behavior, software engineers and hands-on programmers are needed to identify forms of creativity that are tied to the techniques of writing code, designing system architectures, and ensuring interoperability. Multiple methodological skills must be combined to discover what connections exist between technical creativity and cultural practices.

Consider the question of: Where should we look for the next wave of ideas in software development to fuel the nation's economy? We have suggested that creativity and innovation exists in marginal networks outside the mainstream areas of software development.

Corporations are rarely the source of radical new ideas; novel ideas typically come from cottage industries and grass roots organizations. A few examples from the Internet era are blogging, digital music sharing, and Open Source software. These are practices that originated at the margins of society, but have become mainstream, and subsequently adopted by corporations.

Venture capitalists have long understood that the next big thing can come from unlikely corners, so they are willing to take big risks on unknown upstarts because the payoffs can be huge. This observation is consistent with two decades of scholarship in the history of technology, which suggests not only that knowledge often flows from margins to center, but also that the very definition of central or universal knowledge is one that continuously incorporates margins and gives rise to new peripheries. A continuous dynamic emerges: intellectual and popular, mainstream and marginal, core and periphery, interact and shape each other in a historical spiral whose parameters are simultaneously social, political, and technical. Therefore, our hypothesis suggests that more work is needed in the understanding of sub-cultures at the margins (economic, social, and geographic). We should move to investigating margins as not simply "lacking" in the resources of the center but rather as potential sources of radical ideas.

7. REFERENCES

- [1] D. A. MacKenzie and J. Wajcman, *The Social Shaping of Technology, Second Edition*: Open University Press, 1999.
- [2] E. S. Ferguson, "The Mind's Eye: Nonverbal Thought in Technology," *Science*, vol. 197, pp. 827-836, 16 August 1977.
- [3] J. S. Brown and P. Duguid, "Organizing Knowledge," *California Management Review*, vol. 40, pp. 90-111, 1998.
- [4] J. S. Brown and P. Duguid, *The Social Life of Information*: Harvard Business School Press, 2002.
- [5] R. Goldman and R. P. Gabriel, *Innovation Happens Elsewhere: Open Source as a Business Strategy*: Morgan Kaufmann, 2005.
- [6] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*: Cambridge University Press, 1991.
- [7] G. Fischer, "Social Creativity: Turning Barriers into Opportunities for Collaborative Design," in *Eighth Conference on Participatory Design*, Toronto, ON, Canada, 2004, pp. 152-161.
- [8] M. Block, "Library: An Unquiet History: A Review," in *ExLibris*, 2003.
- [9] N. Chanda, *Bound Together: How Traders, Preachers, Adventurers, and Warriors Shaped Globalization*: Yale University Press, 2007.
- [10] "BarCamp wiki - The password is c4mp," 2007.
- [11] "Main Page - BarcampBangalore," 2007.
- [12] P. Long, *Openness, Secrecy, Authorship: Technical Arts and the Culture of Knowledge from Antiquity to the Renaissance*: The Johns Hopkins University Press, 2001.
- [13] L. Daston, "The Moral Economy of Science," in *Osiris, Second Series, Constructing Knowledge in the History of Science*, vol. 10, 1995, pp. 2-24.

- [14] G. Coleman, "The Political Agnosticism of Free and Open Source Software and the Inadvertent Politics of Contrast," *Anthropological Quarterly*, vol. 77, pp. 507-519, Summer 2004.
- [15] Creative Commons, "Creative Commons." vol. 2008.
- [16] EFF, "Electronic Frontier Foundation | Defending Freedom in the Digital World." vol. 2008.
- [17] FSF, "Year end appeal - help protect your freedom! - Free Software Foundation." vol. 2008.

- [18] Linux Kernel Organization Inc., "The Linux Kernel Archives." vol. 2008.
- [19] R. J. Coombe, *The Cultural Life of Intellectual Properties: Authorship, Appropriation, and the Law*: Duke University Press, 1998.
- [20] A. Appadurai, *The Social Life of Things: Commodities in Cultural Perspective*: Cambridge University Press, 1988.

FULL PAPER SUBMISSION 7:

Value-Centered Design for an Aging Populace: Introducing the ETHOS Project

AUTHOR(S):

Shankar, Kalpana
 Connelly, Kay
 Camp, L. Jean
 Lorenzen-Huber, Lesa

ORGANIZATION(S):

Indiana University-Bloomington, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Introduction

The integration of computational tools into everything from our cars and factories to our milk cartons and clothing promises mobile, embedded, universally accessible information technologies for one and all. Pervasive computing uses easily deployed sensors that extract data from their environment, wireless technologies to transmit that data, and then potentially actuating or instantiating other actions by networked technologies according to decisions based on that data. These technologies are increasingly and sometimes invisibly deployed in our lives. Widely deployed systems include temperature sensors, light-controlling movement sensors, 24x 7 CCTV, RFID access devices, clothing tags, and subcutaneous pet identification devices.

One area in which pervasive computing is touted as “gracefully” integrating the human and the computer is in the delivery of health care. The surveillance opportunities afforded by sensor technologies have been pursued in elder care in particular, where monitoring technologies and aggregating databases can help an elder “age in place”, or in a situation that allows for a more “home like” atmosphere. Specifically technologies range from Personal Digital Assistants (PDAs) for monitoring the condition of an ill person to an entire adult residential facility with embedded networked technology.

Home health care is becoming an increasingly prevalent part of American health care, which is not surprising given an aging demographic and improved facilities for maintaining the quality of life as the population ages. Approximately 44 million Americans are caring for an elderly relative, and some estimates by the US Administration on Aging put unpaid health care (for elderly or other incapacitated relatives) at 80% of the national health care budget. Much of this health care is not surprisingly taking place in the home, so it stands to reason that technological supports are looked to for providing some relief and support for the network of informal caregivers.

The potential of ubicomp for home health care is a dual-edge sword that promises great potential for help, but also harm. Privacy advocates and civil libertarians are necessarily concerned about the potential for abuse that ubicomp technologies portend. Even though most designers of ubicomp for health care argue for privacy controls on their technology, their definitions of privacy are often vague at best, and conflicted. This is doubly true for the concept of autonomy, which is the core “driver” of ubicomp technologies in the home health care environment. What are autonomy and privacy, and how do the designers’ definitions of these values become embedded in the technologies they design? And which of the many metaphors of privacy make sense in this environment?

In this paper, we introduce Ethical Technologies in the Homes of Seniors (ETHOS). Project ETHOS, a 3 year National Science Foundation funded project at Indiana University, takes a multidisciplinary, value-centered design approach to answering some these questions. We are currently in year one of the project. Our goals are to:

- Develop a theoretical framework of the privacy concerns of elders, refined through a series of focus groups and explorations in a hands-on “Living Lab”;
- Create tools that can be used by affected populations to make informed decisions about home-based ubicomp;
- Develop a set of designer tools to guide them through a value-centered design approach for ubicomp design
- Create a sensor library and privacy enhancing technologies specific to this use.

Because we intend this project to a model of multidisciplinary I-school research and education, we will expand our presentation to include our methodological approaches and challenges; experiences working with this particular community of users

underaddressed in design - the healthy elderly; and related educational initiatives for I-school graduate and undergraduate research and teaching stemming from this project.

References

United States Agency on Aging (1998). Profile of Older Americans.

Camp, L.J., Huber, L., & Connelly, K. (2006). Beyond consent: implications of ubicomp for privacy. Telecommunications Policy Research Conference 2006. Washington, DC.

Consolvo S. and Towle, J. (2005). Evaluating an ambient display for the home. Proceedings of the CHI '05 Extended Abstracts on Human Factors in Computing Systems (pp. 1304-1307). Portland, OR. ACM.

Hilty, L.M., Som, C., and A. Kohler. (2004). Assessing the human, social, and environmental risks of pervasive computing. Human and Ecological Risk Assessment 10, 853-874.

Mynatt, E., et al (2001). Digital family portraits: supporting peace of mind for extended family members. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (pp. 333-340). Seattle, WA. ACM.

Strickland, L.S. and L.E. Hunt. (2004). Technology, security, and individual privacy: New tools, new threats, and new public perceptions. Journal of the American Society for Information Science and Technology. 56, 221-234.

FULL PAPER SUBMISSION 8:

I-Schools: Mice Roaring or the Future is Now Arriving?

AUTHOR(S):

Sawyer, Steve
Rosenbaum, Howard

ORGANIZATION(S):

Pennsylvania State University, United States of America
Indiana University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Through this essay we develop an argument for the growth of I-Schools by building on the literature and concepts of institutions. We draw on the literatures regarding institutions and institutionalization to help clarify what those involved in the I-School movement might need to do for these scholarly institutions to become a stable and self-replicating form of organization in the university milieu. The literature on institutions and institutionalizing is vast, active, and multi-threaded. In this essay we draw from, but do not summarize or extend this work. We are simply using it to frame our thesis: *what can those involved in I-Schools do to encourage the growth and legitimation of these scholarly institutions?*

On becoming an institution

An institution is a representation of a social order or pattern, continually reproduced, which owes its continued existence to relatively self-activating or automatic social processes (Scott, 1994a, 1995; 2001 DiMaggio & Powell, 1991; Brint & Karabel, 1991; Zucker, 1977, 1983, 1987; Jepperson, 1991). Scott's (1995, 2001) comprehensive definition of an institution consists of cognitive, normative, and regulative structures and activities that provide stability and meaning to social behavior. To Scott institutions as multifaceted systems which incorporate systems of symbols - including cognitive constructions and normative rules - and regulative processes that are carried out through and shape social behavior. In this way, an institution is "a set of roles, graded in authority, that have been embodied in consistent patterns of actions, that have been legitimated and sanctioned by society or segments of that society; whose purpose is to carry out certain activities or prescribed needs of that society or segments of that society." (Mills, 1959)

Something becomes an institution when self-perpetuating internal social patterns reproduce themselves without the need of sustaining action or collective action by its members (Painter, 2002). Instead, routine procedures support and sustain the pattern,

furthering its reproduction, unless collective action blocks or external shocks disrupt the pattern (Jepperson, 1991). Institutions arise due to enacted practices from which evolve stable sets of structures, with formalized rules and laws.

A profession or occupation becomes professionalized or institutionalized when they are believed to require extensive formalized study and a mastery of specialized knowledge; are autonomous; self-restricted and self-regulated; and generally exclusive. A professional institution includes organizations and mechanisms for preserving the knowledge and the practices, enforcing the standards, and educating professionals (Tseng, 1992; Carter, Grebner, Seaman, and Foret, 1990). Drawing from and combining these, we argue here and in other writing that a working definition of a *Scholarly Institution* as one with the following characteristics¹:

1. Automatic: self replicating;
2. Autonomous: self-regulating;
3. Structurally stable: possessing formalized norms and cultures;
4. Visible: having a coherent outward appearance.

The constituent elements and processes that go into forming institutions are a very active area of scholarship (Frumppkin and Kaplan, 2005). Most institutional research has focused on the effect of institutions as independent variables, at the expense of an examination of the determinants and components of institutions and the process by which institutions become constructed remains a “black box” (Zucker, 1988: 104).

One promising approach to understanding the development of institutions is to see it as a structural activity (drawing on Gidden’s theory of structuration) (e.g., DiMaggio & Powell, 1983; DiMaggio & Powell, 1991; Scott, 1994a; Barley and Tolbert, 1997). Framed as a structural process, institutionalization results as orderly patterns or interaction emerge out of loosely organized technical activities serving to infuse a normative order or set of normative values into the organization (Broom and Selznik, 1955). According to tenets of structuration theory, institutions emerge out of a dynamic in which individuals are shaped by institutional forces and then act upon their institutionalized environments to transform institutional arrangements (Barley and Tolbert, 1997). Change comes about in institutional environments when some event or development breaks the patterns established by previously recurrent actions and reflexive behavior of individuals.

What we in I-Schools can do

We argue that becoming a scholarly institution requires takes concerted efforts to become automatic or self replicating; autonomous or self-regulating; structurally stable, by which we mean that they possess formalized norms; and visible: by which we mean that I-Schools have a coherent outward appearance. In the rest of the essay we raise issues and opportunities relative to each of these five characteristics of a scholarly institution.

Pursuing self replication: To achieve this requires both a growth in the number of I-School programs and the development and formalization of shared structures and norms. Several activities suggest this is ongoing, for example there are three mechanisms that bring together the I-School leaders and faculty. The first is the I-Conference. The I-Conference is currently the only community-wide endeavor that engages students, faculty and administrative leadership. The other two – the self-selected and independent I-School caucus and the Computing Research Association (CRA) - sponsored IT Dean’s groups – both provide senior administrators with forums to share structures and develop common norms at an administrative level.

There are at least two aspects of self- replication that deserve additional attention. The first is active encouragement towards faculty and academic leadership of other universities to consider starting or developing I-Schools. The most likely pathway for this is to engage colleagues in scholarly fields such as communications and computing to consider becoming more I-like. There are likely a limited number of universities considering starting from scratch new I-School programs. And, there are relatively few programs whose histories are centered in Library and Information Science who have not begun to transit towards I-School-like programs. Since these two paths are where most of the existing I-Schools drew from, new sources of growth are needed.

The second aspect of self-replication deserving more attention is the development of common structures and norms – elements of a shared culture. Again, there is some progress at an informal level. For example, many of the current I-Schools – and particularly those who have histories in Library and Information Science – have a strong humanistic and informational perspective. Other current I-Schools, however, do not have these norms as deeply-ingrained, yet. The various mechanisms for

¹ See also: **Error! Main Document Only.**Sawyer, S. and Tapia, A. (2007) “From Findings to Theories: Speculating on the Future of Social Informatics” *The Information Society*, 23(4), 263-277.

having I-School faculty and administrative leadership get together provide a means to share and develop common structures. Increasingly, however, this attention should be focusing on ensuring that these structures and norms are part of the undergraduate and graduate training and education. And, for institutional legitimation in the academy, graduate students are often seen as the carriers of common structures and norms.

Pursuing self-regulation: Autonomy in a multi-disciplinary space is one of balance. To maintain multi-disciplinarity demands both development of a common core of knowledge while retaining strong ties to the reference or contributing scholarly disciplines. Without a common core, the space is ephemeral – a collection of local and informal interactions, with no permanence. Without strong connection to reference disciplines, the space becomes an orphan and is too easily ignored.

There are at least three opportunities before I-School members to increase the level of self-regulation. The first is to make explicit a focus on hiring faculty whose intellectual heritage is outside I-Schools along with a commitment to hiring I-School-trained PhD students. This dual strategy is one way to ensure that I-Schools have a good mix of diverse faculty. Second, it is important to develop I-School-trained graduate students and help them secure (and succeed in) faculty positions in the reference or contributing scholarly disciplines. Second, it is important to develop I-School-trained graduate students secure (and succeed) in faculty positions in the reference or contributing scholarly disciplines. Third, there should be more effort to develop future administrative leaders from within I-Schools. Currently this is difficult to do given the limited numbers of I-School faculty and the limited opportunities for I-School faculty to gain administrative experience (there are few departments in the few programs, and few associate dean positions, meaning it is relatively difficult to develop internal-to-I-School candidate pools). One possibility is to begin holding a summer camp for faculty who are interested in pursuing I-School administration. This may also bring in faculty from outside I-Schools and help increase the pool, expand the scale and support self-replication activities.

Pursuing structural stability: Structural stability reflects both a commitment by members to pursue collectively a common set of core issues and to be rewarded for doing so. For I-Schools, this translates into developing the formal and informal norms that recognize and reward faculty for both maintaining a strong connection to research communities that extend beyond I-Schools into reference or contributing academic disciplines while also supporting faculty for engaging more directly (particularly with colleagues at their home institution) in discipline-crossing activities. These norms and incentives should also be reflected in the hiring, tenure, promotion and merit-review processes of faculty

Pursuing visibility: Visibility means having a coherent outward appearance. This has been the focus of the I-School caucus and academic leadership, and should be a continued focus. There is also evidence that faculty are beginning to help make I-Schools more visible. For example, the I-Schools and I-School faculty are very involved in the human-computer interaction (HCI), participatory design, computer-supported cooperative work and social informatics research communities. This involvement leads to where others in these communities, and particularly younger faculty and graduate students, see I-Schools as a viable, if not preferred, choice for an academic home.

The evidence suggests that I-School faculty and administrative leadership are making steps towards becoming a scholarly institution. The need for collective action to be focused and to pursue the efforts needed to ensure institutionalization benefit all of us. These actions also demand our time and attention.

References

- Barley, S. and Tolbert, P. (1997) "Institutionalization and Structuration: Studying the links between institutions and actions," *Organization Studies*, 18, 93-117.
- Broom, L. and Selznick, P. (1955) *Sociology*, Row & Peterson, Evanston, Illinois.
- Tseng, J. (1992). Interpreting as an Emerging Profession in Taiwan -- A Sociological Model. Unpublished manuscript, Fu Jen Catholic University, Taiwan.
- Brint, S. & Karabel, J. (1991). Institutional Origins and Transformations: The Case of American Community Colleges, In *The New Institutionalism in Organizational Analysis*, ed. Walter W. Powell and Paul J. DiMaggio, 337-360. Chicago: University of Chicago Press.

- Carter, M., Grebner, F., Seaman, J. & Foret, C. (1990). Occupation to Profession Continuum --Status and Future of AAHPERD". *Journal of Physical Education, Recreation and Dance*, 61(3), 1-14.
- DiMaggio, P. & Powell, W. (1991). Introduction, In *The New Institutionalism in Organizational Analysis*. ed. W. Powell and P. DiMaggio, University of Chicago Press, Chicago, 1-38.
- DiMaggio, P. & Powell, W. (1991) The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields, In *The New Institutionalism in Organizational Analysis*, W. Powell and P. DiMaggio, University of Chicago Press, Chicago, 63-82.
- Jepperson, R. (1991) Institutions, Institutional effects, and Institutionalization, In *The New Institutionalism in Organizational Analysis*, W. Powell and P. DiMaggio (eds.), University of Chicago Press, Chicago, 143-163.
- Mills, C. (1959). *The sociological imagination*, Oxford University Press, London.
- Painter, C. (2002). Early Leader Effects on the Process of Institutionalization Through Cultural Embedding: The Cases of William J. Donovan, Allen W. Dulles, and J. Edgar Hoover, Unpublished Dissertation. Virginia Polytechnic Institute and State University.
- Scott, W. (1994). Institutions and Organizations: Toward a Theoretical Synthesis, In *Institutional Environments and Organizations: Structural Complexity and Individualism*. W. R. Scott, J. Meyer, et. al., 55-80. Thousand Oaks, CA: Sage Publications.
- Scott, W. (1995). Introduction: Institutional Theory and Organizations, In *The Institutional Construction of Organizations: International and Longitudinal Studies*. ed. W. Scott and S. Christensen, xi-xxiii. Thousand Oaks, CA: Sage Publications.
- Scott, W. (2001). *Institutions and Organizations*. Thousand Oaks, CA: Sage.
- Zucker, L. (1977). The Role of Institutionalization in Cultural Persistence, *American Sociological Review*, 42(5), 726-743.
- Zucker. (1983). Organizations as Institutions, in Samuel B. Bacharach ed. *Research in the Sociology of Organizations*, 2, 1-47. Greenwich, CT: JAI Press.
- Zucker. (1987). Institutional Theories of Organizations, *Annual Review of Sociology*, 13, 443-464.
- Zucker, L. (1988). Where Do Institutional Patterns Come From?, In *Institutional Patterns and Organizations*. ed. L. Zucker, 21-43. Cambridge, Mass: Ballinger.

FULL PAPER SUBMISSION 9:

Understanding delicious Tag Choice Using Simulations

AUTHOR(S):

Wash, Rick
Rader, Emilee

ORGANIZATION(S):

University of Michigan, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Understanding del.icio.us Tag Choice Using Simulations

Paper Submission

User-contributed metadata, also known as *tagging*, is increasingly receiving attention as a tool for digital information management. Collaborative tagging systems such as del.icio.us and citeulike.org publicly expose individual users' associations between content items and tags, thereby providing visibility into words others have used to tag similar items. Grudin (2006) suggests that collaborative tagging can be a low-effort solution for shared or group information management, because it does not require that users try to conform to a controlled vocabulary or organization scheme.

In this paper, we focus on the social bookmarking website del.icio.us as a case study of a collaborative tagging system supporting both personal and shared information management. del.icio.us is an online application that allows users to save and tag their own web bookmarks so they are accessible from any networked computer. del.icio.us has recently received attention in the research literature as the canonical example of a collaborative tagging system for information management (Golder and Huberman, 2006; Halpin et al., 2007). We are studying how users choose tags, and hope to apply this knowledge to improve information management interfaces.

Wash and Rader (2007) argue that the usefulness of del.icio.us depends critically on how users choose tags. Golder and Huberman (2006) argue that users' tag choices are not random; instead, consensus seems to emerge for which tags best represent a given web page. They show that web pages bookmarked in del.icio.us demonstrate a stable frequency distribution following a power-law pattern in which the same few tags are chosen by many users, while most other tags are selected by only a few users. The "long tail" of the power-law distribution includes hundreds more tags used by only one or two people. Golder and Huberman speculate that this pattern might be due to users imitating each others' tag choices; in other words, when a user bookmarks a web page in del.icio.us, their tag choices might be influenced by tags that had been previously applied to that web page by other users.

However, it is reasonable to assume that there might be other sources of influence on users' tag choices having to do with personal information management goals. For example, a user interested in del.icio.us only for organizing and re-finding their own bookmarks might strive for consistency within their own "controlled vocabulary", to maintain a shorter list of tags (Wash and Rader, 2007). Or, users might desire to expend as little effort as possible when choosing tags, and simply select the tags the system recommends when they create a new bookmark.

To better understand tag choices, we developed a computational simulation of tagging behavior on del.icio.us. This simulation enabled us to "assume" different strategies for choosing tags and see what the results would look like on del.icio.us. We ran the simulation for four different tagging strategies and compared the resulting tag frequency distributions to the distribution observed on del.icio.us. We then tried to determine if the proposed

tagging strategies are consistent with the empirically observed data. We tested the following strategies:

Zipf's Law: Zipf's law states that word frequency in most written works follows a powerlaw distribution. Therefore, del.icio.us users might naturally choose their words from a powerlaw distribution.(Newman, 2005)

Imitation: Users might imitate previous users' tag choices. This was described by Golder and Huberman (2006)

User-based: Users might favor tags that they had used previously. This was described by Wash and Rader (2007)

Recommended: Users might prefer to click on the tags that are programatically recommended by del.icio.us.

The latter three are the same strategies studied in a large-scale statistical analysis by Rader and Wash (2007).

Empirical Results

In order to simulate tag choices on del.icio.us, and to compare the simulations back to del.icio.us, we found it important to rigorously estimate the distribution of tags on del.icio.us. We collected the tagging history of approximately 12,000 URLs that had been bookmarked on del.icio.us. We also collected the tagging history of approximately 12,000 users from del.icio.us who had bookmarked any of 30 specific websites. We used this data to estimate two distributions: 1) The distribution of tag frequency — for a given website bookmarked, how frequently was each tag applied to it? and 2) The number of tags chosen by a user — when a user is bookmarking a website, how many tags will he or she apply?

Unfortunately, there is no known method for determining which theoretical distribution generated a set of data. However, we can fit the data to multiple families of distributions and see which distribution fits “best.” “Best” here is a statistical determination (Clauset et al., 2007). We fit the data from each site to seven different discrete probability distribution families (discrete powerlaw, negative binomial, binomial, discrete lognormal, discrete exponential, poisson, and geometric), estimating parameters with maximum likelihood estimation. We then used a non-nested Kolmogorov-Smirnov test to conduct pairwise comparisons between these distributions.

The discrete powerlaw distribution fit the empirically observed tag distributions better than any of the other 6 distributions we tested. The fitted distribution had an average exponent (α) of 1.92 ± 0.40 . This is a rather low exponent for a powerlaw distribution, and indicates that the “long tail” of tags is very long and heavy. This low exponent also has another important implication. Newman (2005) explains that powerlaw distributions with an exponent less than 2 have an infinite (or undefined) mean. Therefore, estimates of a “mean” or average tag are undefined, and any inferential statistics based on the mean of the tag distribution cannot be used.

The number of tags chosen by a user fit a discrete log normal distribution better than the other 6 distributions we tested. The fitted distributions had a mean log value of 0.82 ± 0.45 and a standard deviation log value of 0.73 ± 0.19 . On average a user of del.icio.us will choose 2.51 tags (with a standard deviation of 1.42) when bookmarking a website.

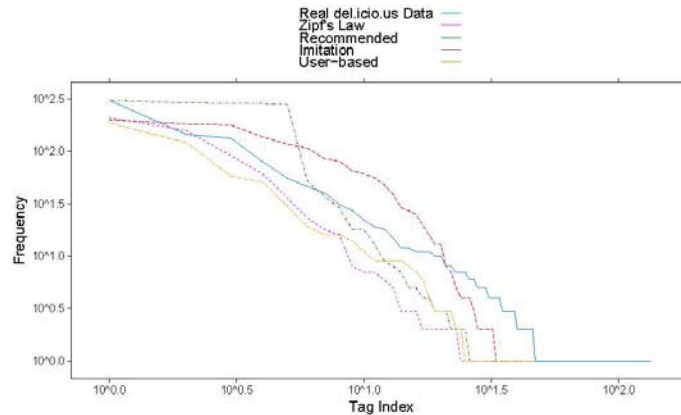


Figure 1: Tag distributions for four simulated runs, matched with a real empirically-observed tag distribution (on a log-log scale)

Simulation Results

One of the benefits of simulation is that the development process forced us to be very explicit about what information users would need to follow a hypothesized strategy. Golder and Huberman (2006) suggest that the powerlaw distribution of tags for a given website could arise from users intentionally imitating previous users tag choices. They cite the networks literature (Barabasi and Albert, 1999; Newman, 2005) for evidence that powerlaw distributions can arise from such path-dependent choices. When trying to replicate these decisions for our simulation, we found that this only works if a user chooses tags from the empirical distribution *at the time of decision*. This means if tag 'A' has been used twice as much as tag 'B', then tag 'A' need to be chosen with twice the probability of tag 'B'. This is a very high information requirement for users – they must know the exact proportions of existing tags to choose appropriately. Del.icio.us does not make this information available; however we assume this knowledge for our simulations with the **imitate** strategy.

To compare the four tag choice strategies, we simulated 120 “websites” for each of the four strategies. Each of the 120 websites was paired with one of 30 real websites, and the number of users was chosen to match the real website. Figure 1 shows the tag distribution (on a log-log plot) for all four of the strategies on one of these websites. Also on the graph is the empirically observed tag distribution from the paired website. The non-powerlaw nature of the **recommended** strategy stands out, with a small number of roughly equally likely tags (the recommended tags) and then a sharp drop in probability for the other tags.

For each simulated website, we fit the simulated tag distribution to a discrete power-law distribution. We used this fit for two comparisons. First we computed the difference between the powerlaw exponent in the simulated distribution and the exponent in the real distribution. The first column (Delta) of Table 1 shows the average difference in exponents.

	<i>Delta</i>	<i>KS</i>
Zipf's Law	0.23	0.08
User	0.19	0.08
Path-dependent	0.32	0.14
Recommended	0.34	0.22

Table 1: 1) Average difference between simulated powerlaw exponent and real powerlaw exponent, and 2) Average Kolmogorov-Smirnov goodness-of-fit statistic

Second, we conducted a Kolmogorov-Smirnov goodness-of-fit test to see how well the simulated distribution fit a powerlaw. A KS statistic of 0 means that the distribution is identical to a powerlaw, and higher numbers indicate greater deviation from a powerlaw. The second column (KS) of Table 1 indicates the average KS statistic for each of the four strategies. As a comparison, the empirical data from del.icio.us fits a powerlaw with an average KS statistic of 0.067. Neither the **imitation** nor **recommended** strategies fit a powerlaw very well. The distributions from **Zipf's Law** and **User-based** strategies fit as well as the real data from del.icio.us.

Conclusions

We compared four possible strategies that users might use to choose tags on del.icio.us. Our simulations indicate that choosing the recommended tags would result in a skewed distribution that fits a powerlaw distribution less than the empirically observed distribution. Additionally, the hypothesis that users choose tags by imitating other users, with previous tag choices influencing the current choice, is also unlikely. The information requirements that would be necessary for users to choose tags this way are large, and even if they did, the resulting distributions are different than the empirically observed tag distributions. Neither the Zipf's law hypothesis nor the user-based hypothesis can be ruled out based on our simulations. These results corroborate the work of Rader and Wash (2007).

References

- Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286, 1999. URL <http://arxiv.org/abs/cond-mat/9910332>.
- Aaron Clauset, Cosma Rohilla Shalizi, and Mark E. J. Newman. Power-law distributions in empirical data. Preprint, Jun 2007. URL <http://arxiv.org/abs/0706.1062v1>.
- Scott Golder and Bernardo A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.
- Jonathan Grudin. Enterprise knowledge management and emerging technologies. In *HICSS '06*, 4-7 January 2006.
- Harry Halpin, Valentin Robu, and Hana Shepherd. The complex dynamics of collaborative tagging. In *WWW '07*, 2007.
- M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323–351, 2005.

Understanding del.icio.us Tag Choice Using Simulations

Rick Wash

School of Information
University of Michigan
rwash@umich.edu

Emilee Rader

School of Information
University of Michigan
ejrader@umich.edu

ABSTRACT

Understanding how users choose tags can help researchers better understand how tagging systems can be used and how to design better tagging systems for the future. We developed a simulation of del.icio.us, a popular social bookmarking tool, that allowed us to simulate users choosing tags using one of four possible strategies for tag choice found in the literature. We then compared the resulting tag choices with empirical data retrieved from del.icio.us to determine which tag choice strategies would result in choices most similar to those seen in the real world. We were able to rule out three of the strategies as unlikely to be the primary means by which tags are chosen on del.icio.us.

Topics

Information Organization

Keywords

tagging, del.icio.us, social bookmarking, simulation

1. INTRODUCTION

User-contributed metadata, also known as *tagging*, is increasingly receiving attention as a tool for digital information management. Collaborative tagging systems such as del.icio.us and citeulike.org publicly expose individual users' associations between content items and tags, thereby providing visibility into words others have used to tag similar items. Grudin [8] suggests that collaborative tagging can be a low-effort solution for shared or

group information management, because it does not require that users try to conform to a controlled vocabulary or organization scheme.

In this paper, we focus on the social bookmarking website del.icio.us as a case study of a collaborative tagging system supporting both personal and shared information management. del.icio.us is an online application that allows users to save and tag their own web bookmarks so they are accessible from any networked computer. del.icio.us has recently received attention in the research literature as the canonical example of a collaborative tagging system for information management [7, 9]. We are studying how users choose tags, and hope to apply this knowledge to improve information management interfaces.

Wash and Rader [12] argue that the usefulness of del.icio.us depends critically on how users choose tags. Golder and Huberman [7] argue that users' tag choices are not random; instead, consensus seems to emerge for which tags best represent a given web page. They show that web pages bookmarked in del.icio.us demonstrate a stable frequency distribution following a power-law pattern in which the same few tags are chosen by many users, while most other tags are selected by only a few users. The "long tail" of the power-law distribution includes hundreds more tags used by only one or two people. Golder and Huberman speculate that this pattern might be due to users imitating each others' tag choices; in other words, when a user bookmarks a web page in del.icio.us, their tag choices might be influenced by tags that had been previously applied to that web page by other users.

However, it is reasonable to assume that there might be other sources of influence on users' tag choices having to do with personal information management goals. For example, a user interested in del.icio.us only for organizing and re-finding their own bookmarks might strive for consistency within their own "controlled vocabulary", to maintain a shorter list of tags [12]. Or, users might desire to expend as little effort as possible when choosing tags,

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a license for the iSchools to post and disseminate their papers on this site and on a CD. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse this work. The authors are solely responsible for their paper's content. Our thanks to the Association for Computing Machinery for permission to adapt and use their template for the iSchools 2008 Conference.

and simply select the tags the system recommends when they create a new bookmark.

The research literature contains multiple competing explanations for the patterns of tags that appear on del.icio.us. All of these explanations initially seem like reasonable explanations of the way that del.icio.us users choose tags. But they cannot all be right simultaneously. It would be nice to actually have collections of real people specifically use one or more of the strategies suggested by the literature on delicious. We could then see if those tag choices resulted in patterns of tags similar to those found on del.icio.us. Unfortunately, this technique would be prohibitively costly. Instead, we developed a computational simulation of users doing this. This simulation allows us to control exactly what strategy the simulated users used to choose tags. Such a simulation cannot tell us which strategy the real users of del.icio.us used; it can only tell us which strategies are likely to result in patterns of tags that are similar to those observed on del.icio.us. In other words, the simulation cannot confidently identify how users chose tags, but it can be used to rule out explanations that are unlikely to generate the observed patterns of tags.

We ran the simulation for four different tag selection strategies and compared the results with the tag frequency distribution observed on del.icio.us. We then used statistical techniques to determine whether the simulated tag selection strategies were consistent with the empirically observed data. We tested the following strategies:

Zipf's Law: Zipf's law states that word frequency in most written works follows a powerlaw distribution. Therefore, del.icio.us users might naturally choose their words from a powerlaw distribution.[10]

Imitation: Users might imitate previous users' tag choices. This was described by Golder and Huberman [7]

User-based: Users might favor tags that they had used previously. This was described by Wash and Rader [12]

Recommended: Users might prefer to click on the tags that are programatically recommended by del.icio.us.

The latter three are the same strategies studied in a large-scale statistical analysis by Rader and Wash [11].

The goal of the research described in this paper was to detect similarities between the tag frequency distributions

produced by our four simulated strategies, and the actual frequency distribution of tags on a random sample of websites bookmarked in del.icio.us, henceforth called the *empirical* distribution. In this way, we hoped to identify which strategies were capable of producing distributions similar to the empirical distribution and which were not. Strategies producing similar distributions can be considered plausible, in the sense that they could possibly have given rise to the empirical distribution. Strategies producing distributions that are dramatically different from the empirical distribution can be ruled out as unlikely to have been used widely on del.icio.us

We began by downloading and parsing the entire bookmark and tag histories for approximately 12,000 different websites in del.icio.us. We generated the tag frequency distribution for a randomly selected subset of those websites, and compared it against seven known distribution types (powerlaw, log-normal, geometric, etc.), to determine the best fit. Knowing the family of distributions to which the del.icio.us tag frequency distribution belongs allowed us to statistically determine whether the distributions produced by our simulation were of the same type.

Next, we developed a simulation of users' tag choices, details of which will be provided later in the paper. The simulation was instructed to follow each one of the four strategies listed above, in turn. For each strategy, 30 different websites were simulated; for each website users' simulated tag selections were recorded and statistically compared with the empirical distribution.

We were primarily interested in looking at two specific patterns of tags. First, as many people have pointed out, the tags associated with a specific website in del.icio.us tend to follow a powerlaw distribution [7]. Tag choice strategies that do not result in a powerlaw-shaped distribution of tags are unlikely to have been used by the users of del.icio.us. This powerlaw distribution seems to be an important property of tags on del.icio.us [7]. Second, we have looked at the average inter-user agreement between users who bookmark the same website. Furnas et al. [6] have found that typically two individuals will choose the same word to describe an object less than 20% of the time. On del.icio.us, this number is closer to 15% of the time. Tag choice strategies that result in many users agreeing with each other are also unlikely to have been used on del.icio.us. Inter-user agreement is a metric that is sufficiently different than the powerlaw distribution of tags, and is a good complementary metric for characterizing a set of tag choices.

In this paper, we first report the results from our analysis of the empirical data downloaded from del.icio.us. The

results from this analysis were used in the simulation to make the simulation more realistic. We then describe the simulation in detail, and report the results of the analysis comparing the empirical and simulated data. We conclude with a discussion of the implications of this work.

2. ANALYSIS OF EMPIRICAL DATA

Over two weeks in January 2007, we downloaded the entire bookmark and tag history for approximately 20,000 different webpages in del.icio.us. The webpages were chosen by periodically sampling the “recently posted” and “popular” del.icio.us pages. We randomly chose 30 webpages from our sample that had been bookmarked by at least 100 users. Then, in June 2007 we downloaded the complete public bookmark histories for all of the approximately 12,000 users who had ever bookmarked any of these 30 webpages. In other words, our dataset contains the complete tag histories for 30 webpages bookmarked in del.icio.us, as well as tag histories for all users who ever bookmarked any of those 30 webpages.

We used this data to estimate two distributions: 1) The distribution of tag frequency — for a given website bookmarked, how frequently was each tag applied to it? and 2) The number of tags chosen by a user — when a user is bookmarking a website, how many tags will he or she apply?

We fit the data to multiple families of distributions and see which distribution fits “best.” “Best” here is a statistical determination [4]. We fit the data from each site to seven different discrete probability distribution families (discrete powerlaw, negative binomial, binomial, discrete lognormal, discrete exponential, poisson, and geometric), estimating parameters with maximum likelihood estimation. We then used a non-nested Kolmogorov-Smirnov test to conduct pairwise comparisons between these distributions.

The discrete powerlaw distribution fit the empirically observed tag distributions better than any of the other 6 distributions we tested. The fitted distribution had an average exponent (α) of 1.92 ± 0.40 . This is a rather low exponent for a powerlaw distribution, and indicates that the “long tail” of tags is very long and heavy. This low exponent also has another important implication. Newman [10] explains that powerlaw distributions with an exponent less than 2 have an infinite (or undefined) mean. Therefore, estimates of a “mean” or average tag are undefined, and any inferential statistics based on the mean of the tag distribution cannot be used.

The number of tags chosen by a user fit a discrete log

normal distribution better than the other 6 distributions we tested. The fitted distributions had a mean log value of 0.82 ± 0.45 and a standard deviation log value of 0.73 ± 0.19 . On average a user of del.icio.us will choose 2.51 tags (with a standard deviation of 1.42) when bookmarking a website.

Inter-user Agreement It has long been accepted that people use language imprecisely, and meaning is negotiated on the fly during conversation [3]. This imprecision is evident not only in communication, but also when people are asked to create keywords for recipes and names for common editing operations [6] and when user-generated index terms are compared with Library of Congress subject headings [5]. In fact, the probability that two people will generate the same label for the same object, called the “vocabulary problem,” is widely held to be less than 20% [6, 2]. When a user wants to take advantage of the collective properties of social bookmarking by browsing or searching on tags, the vocabulary problem becomes apparent. If users are unlikely to choose the same tags to represent the same topics, such diversity would decrease search precision. When a given tag is applied to bookmarks in an inconsistent manner by many users, more variability exists in the content returned when a user searches with that tag. The desired bookmark may be returned, but there would be too much other “noise” in the results for it to be noticed.

To measure the extent of the vocabulary problem on del.icio.us, we calculated the average inter-user agreement for a sample of 200 users from each of the 30 websites that we had full user data for. On average, users who bookmarked these websites chose the same tag for the website only $14\% \pm 5\%$ of the time. This percentage is low, indicating a fair bit of disagreement between users, though it is higher than the 8% reported by [6] for their text-editing operations dataset. Figure 1 shows a histogram of the inter-user agreement values for the 30 websites.

3. SIMULATING TAG CHOICES

To compare the four tag choice strategies, we simulated 120 websites for each of the four strategies. Each of the 120 websites was paired with one of 30 real websites randomly selected from our sample downloaded from del.icio.us, and the number of users for each simulated website was chosen to match the real website.

Each simulated website was assigned to have the same number of users as its matched real website, and each simulated user was matched with a real user who bookmarked that website. In essence, we are simulating what would happen if the same set of users bookmarked the real

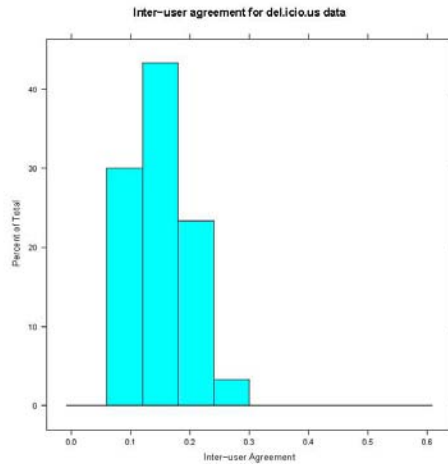


Figure 1: Histogram of the Average Inter-user Agreement for 30 websites bookmarked on del.icio.us

website, but chose their tags according to one of our four hypothesized strategies (and bookmarked it in a random order). To simulate a user choosing tags for that website, two choices have to be made. First, the simulator chooses how many tags that user will apply to the website. Second, the simulator chooses which specific tags would be applied.

As we found above for del.icio.us, the number of tags applied by a user tends to be distributed according to a discrete log normal distribution. For each simulated website, the simulation randomly chose a set of parameters (mean log, and standard deviation of the log) for a discrete log normal probability distribution to match the distribution of parameter values we found empirically on del.icio.us. Once a set of parameters was chosen, this specifies an exact probability distribution. For each simulated user, the simulator would then choose a random number from this probability distribution, and that number would be the number of tags that simulated user would choose.

We found that that tags applied to a given website on del.icio.us are distributed according to a discrete powerlaw distribution. In a manner similar to that used for the number of tags, the simulator chooses a parameter (α) for a discrete powerlaw distribution. This distribution then will serve as the base distribution that specific tags are drawn from. The α parameter is chosen not to directly match the value observed above (1.92), but to

average 2.5. This was done because any single user's tags cannot repeat, and this lack of repetition (i.e., sampling without replacement) forces the user to choose more tags in the 'tail' of the distribution than they would if tag choices were truly independent. If I choose three tags, they cannot all be the most popular tag. Through experimentation, we found that having users choose tags from a powerlaw distribution with an α about 2.5 tends to yield a site-wide powerlaw distribution for tags that approximately match the observed distributions.

Each number from this distribution is then mapped onto a specific tag, according to its rank in the frequency distribution. The tags from the matched real site are ordered from most-frequently used to least-frequently used, with ties being broken randomly. A 1 from the random number generator is then mapped onto the most-frequently-used tag, 2 onto the second most frequently used tag, and so on. Any numbers larger than the number of tags on the matched site are left as numbers.

For each user, the specific tags that they choose will depend on which of the four strategies they are assigned to use. The only difference between these four strategies is in specific tag choice; all other decisions (number of users, number of tags per user, etc.) are identical.

Zipf's Law The simplest strategy is to follow Zipf's law, and choose tags directly from the base powerlaw distribution. The simulator continually chooses random numbers from the base powerlaw distribution until it has the required number of unique numbers. These numbers are then mapped onto tags as described above.

Imitation For users to imitate previous users' tag choices, it is necessary for those previous users to exist. The first few users who bookmark a website will have no one to imitate. To handle this, the first 20 users will draw as described above for Zipf's law and serve as 'seeders.' All users after the first 20 who use this strategy will choose a tag from the current empirical distribution of tags for his simulated website. This means that if there are two tags, 'A' and 'B', and 'A' has been used twice previously and 'B' only once, then tag 'A' is chosen with probability $\frac{2}{3}$ and tag 'B' is chosen with probability $\frac{1}{3}$. However, to ensure growth of the vocabulary beyond that used by the initial 20 seeders, each tag choice has a 10% probability of choosing a new, previously unused tag. This probability was chosen to match the average empirically observed probability from the del.icio.us data. The average website in our sample from del.icio.us has a new tag probability of $10.5\% \pm 8.3\%$.

User-based Not all users will want to apply tags that they have used before. As such, when simulating the **user-based** strategy, the simulated users had a 50/50 chance of choosing tags according to Zipf's law, and of choosing tags they had used before. When choosing tags they had used before, the simulator computes the overlap (set intersection) between the tags the user had ever used and the tags that were ever applied to the matched site. It then randomly chooses among the tags in this overlap set. If that is not enough tags, then additional tags are chosen randomly from the base powerlaw distribution.

Recommended Del.icio.us does not make their algorithm for choosing which tags to recommend public. As such, we could not directly simulate users choosing from del.icio.us's recommended tags. We did, however, create a simple approximation. We proposed that the tagging system could simply recommend the N most popular tags for that website. Then users could randomly choose among those N tags.

If del.icio.us's recommendation algorithm is solely based on the frequency of applying tags to that website, then this approximation is a reasonable proxy for del.icio.us's recommendation algorithm. If the real algorithm includes other data (such as the tags a user has used, or the popularity of tags across the site) then this simulated strategy will not be a good proxy for the del.icio.us algorithm.

To simulate users choosing tags with the **recommended** strategy, we first create 20 'seeders' in the same way we did for the **imitation** strategy simulation. All of the remaining users then are simulated to have been presented with $N = 5$ 'recommended' tags (the 5 most popular tags at that point) and then randomly choose between these recommended tags. If they need to apply more than 5 tags, then the remaining tags are chosen randomly from the base powerlaw distribution.

4. SIMULATION RESULTS

One of the benefits of simulation is that the development process forced us to be very explicit about what information users would need to follow a hypothesized strategy. Golder and Huberman [7] suggest that the powerlaw distribution of tags for a given website could arise from users intentionally imitating previous users tag choices. They cite the networks literature [1, 10] for evidence that powerlaw distributions can arise from such path-dependent choices. When trying to replicate these decisions for our simulation, we found that this only works if a user chooses tags from the empirical distribution *at the time of decision*. This means if tag 'A' has been used twice as much as tag 'B', then tag 'A' need to be chosen with twice

Table 1: 1) Average difference between simulated powerlaw exponent and real powerlaw exponent, and 2) Average Kolmogorov-Smirnov goodness-of-fit statistic

	<i>Delta</i>	<i>KS</i>
Real World	X	0.07
Zipf's Law	0.23	0.08
User	0.19	0.08
Imitation	0.32	0.14
Recommended	0.34	0.22

the probability of tag 'B'. This is a very high information requirement for users – they must know the exact proportions of existing tags to choose appropriately. Del.icio.us does not make this information easily available; however we assume this knowledge for our simulations with the **imitate** strategy.

To compare the four tag choice strategies, we simulated 120 "websites" for each of the four strategies. Each of the 120 websites was paired with one of 30 real websites, and the number of users was chosen to match the real website. Figure 2 shows the tag distribution (on a log-log plot) for all four of the strategies on one of these websites. Also on the graph is the empirically observed tag distribution from the paired website. The non-powerlaw nature of the **recommended** strategy stands out, with a small number of roughly equally likely tags (the recommended tags) and then a sharp drop in probability for the other tags.

For each simulated website, we fit the simulated tag distribution to a discrete powerlaw distribution. We used this fit for two comparisons. First we computed the difference between the powerlaw exponent in the simulated distribution and the exponent in the real distribution. The first column (Delta) of Table 1 shows the average difference in exponents. Second, we conducted a Kolmogorov-Smirnov goodness-of-fit test to see how well the simulated distribution fit a powerlaw. A KS statistic of 0 means that the distribution is identical to a powerlaw, and higher numbers indicate greater deviation from a powerlaw. The second column (KS) of Table 1 indicates the average KS statistic for each of the four strategies. As a comparison, the empirical data from del.icio.us fits a powerlaw with an average KS statistic of 0.067. Neither the **imitation** nor **recommended** strategies fit a powerlaw very well. The distributions from **Zipf's Law** and **User-based** strategies fit as well as the real data from del.icio.us.

Inter-user Agreement In addition to fitting the tag distribution, we also calculated the average inter-user agreement between users of each of the 480 simulated websites. Figure 3 shows the distribution of average inter-

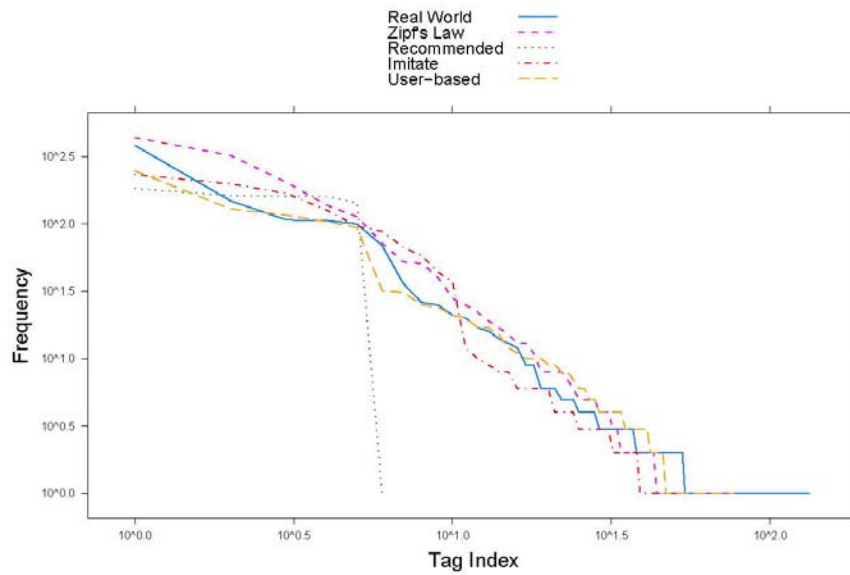


Figure 2: Tag distributions for four simulated runs, matched with a real empirically-observed tag distribution (on a log-log scale)

Table 2: Average Inter-user Agreement for each of the four simulated tagging strategies

	<i>Mean IUA</i>	σ^2
Real World	0.144	0.051
Zipf's Law	0.373	0.074
User	0.182	0.052
Imitation	0.184	0.056
Recommended	0.317	0.088

user agreements for each of the four strategies. Table 2 provides the mean inter-user agreement value for each of the four strategies (along with their standard deviation), and a comparison point with the real-world data from del.icio.us. The mean inter-user agreement for the **user-based** strategy and the **imitation** strategy are statistically indistinguishable. All other pairwise comparisons between inter-user agreement values are statistically significant at the 10% level.¹

Our simulations resulted in users agreeing with each other much more often when using the **Zipf's law** strategy or the **recommended** strategy. This indicates to us that these strategies are unlikely to have been used by the users who bookmarked these 30 websites on del.icio.us. However, the inter-user agreement tests cannot rule out either the **user-based** strategy or the **imitation** strategy for tag choices, because they resulted in similar levels of inter-user agreement.

5. DISCUSSION

We compared four possible strategies that users might use to choose tags on del.icio.us. Our simulations indicate that choosing the recommended tags would result in a skewed distribution that fits a powerlaw distribution less than the empirically observed distribution. Choosing the recommended tags will also cause higher levels of inter-user agreement than are empirically observed. Both of these findings are a direct result of the fact that in our simulation users choose uniformly at random among the recommended tags. They do this because they have no better way of determining which recommended tags are appropriate. On del.icio.us, it is unclear how users choose among the tags that del.icio.us recommends. If they just 'click a couple' of tags, then del.icio.us would end up with skewed tag distributions and high inter-user agreement. Since we did not find that on del.icio.us, if users do utilize

the recommended tags our simulations suggest that they do so more intentionally than 'just clicking a few' of the recommended tags.

The hypothesis that users choose tags by imitating other users, with previous tag choices influencing the current choice, is also unlikely to be what del.icio.us users are really doing. The information requirements that would be necessary for users to choose tags this way are large; users need to know both which words were previously applied as tags and how frequently they were applied to use this hypothesized strategy. But even if users could easily get the proper information, the resulting tag distributions from the simulations are skewed and have a shorter tail than the tag distributions we observed on del.icio.us. This suggests that users are driven by something more than just imitating past users.

We also believe that the Zipf's law hypothesis is unlikely, as it results in dramatically higher inter-user agreement than we observed on del.icio.us. This high inter-user agreement is because every user who bookmarks a website with this strategy is choosing tags from the same power-law distribution — the tag that is most likely to be chosen is the same for all users. This suggests that users are not all choosing tags from the same distribution — at the very least they have individualized distributions of words to choose from.

We cannot rule out the user-based hypothesis based on our simulations. We found that when our simulated users choose tags using the user-based strategy, the resulting tag distribution is as close to a powerlaw distribution as our empirical data from del.icio.us, and the inter-user agreement is approximately similar to the level of inter-user agreement in our sample of del.icio.us. Choosing tags from the user's set of tags results in a similar level of inter-user agreement and a powerlaw tag distribution. This suggests that each user has his or her own way of choosing tags, and that individual and idiosyncratic way of choosing tags is a major influence on tag choices. However, this research is not able to distinguish what some of the different idiosyncratic strategies are. It is clear, though, that future systems that support collaborative tagging will need to be sufficiently flexible to allow each user to choose their own way of determining which tags to use.

Since this is a simulation study, we can only compare aggregate measures of tag choices like the shape of the resulting tag distribution. This study cannot look at individual tag decisions to determine which strategy was used for that specific tag choice, as the same strategy might not be used across all users, or even all the tag choices of an

¹Comparisons were done with a series of t-tests, and used a Bonferroni correction to adjust the p-values for multiple tests. All remaining comparisons were statistically significant at the 0.1% level except the two: the comparison between the real world data and the **user-based** strategy, and the comparison between the real world data and the **imitation** strategy.

individual user. In another part of this project, we [11] use a logistic regression to attempt to determine which of these four strategies was in use for individual tagging decisions on del.icio.us. Fortunately, the results from that work corroborate the results here.

REFERENCES

- [1] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286, 1999.
- [2] M. J. Bates. Indexing and access for digital libraries and the internet: Human, database, and domain factors. *Journal of the American Society for Information Science*, 49(13):1185–1205, 1998. Bates.JA-SIS1998.
- [3] H. Clark. *Using Language*. Cambridge University Press, Cambridge, England, UK, 1996.
- [4] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. Preprint, Jun 2007.
- [5] L. Y. Collantes. Degree of agreement in naming objects and concepts for information retrieval. *Journal of the American Society for Information Science*, 46(2):116–132, 1995.
- [6] G. Furnas, T. Landauer, L. Gomez, and S. Dumais. Statistical semantics: Analysis of the potential performance of key-word information systems. *The Bell System Technical Journal*, 62(6):1753–1806, 1983.
- [7] S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.
- [8] J. Grudin. Enterprise knowledge management and emerging technologies. In *HICSS '06*, 4-7 January 2006.
- [9] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *WWW '07*, 2007.
- [10] M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323–351, 2005.
- [11] E. Rader and R. Wash. Collaborative tagging and information management: Influences on tag choices in del.icio.us. Working Paper, University of Michigan., September 2007.
- [12] R. Wash and E. Rader. Public bookmarks and private benefits: An analysis of incentives in social computing. In *ASIS&T Annual Meeting*, 2007.

FULL PAPER SUBMISSION 10:

Whose data do you trust? Integrity issues in the preservation of scientific data

AUTHOR(S):

Mayernik, Matthew S.

Wallis, Jillian C

Pepe, Alberto

Borgman, Christine L

ORGANIZATION(S):

University of California, Los Angeles, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Whose data do you trust? Integrity issues in the preservation of scientific data

Matthew S Mayernik¹, Jillian C Wallis², Alberto Pepe¹, & Christine L Borgman¹

¹ GSE&IS, University of California, Los Angeles

² Center for Embedded Networked Sensing

Preserving digital information is a central concern to the design of information infrastructure. Digital information takes many forms, and one of increasing importance is scientific data. Data are now being seen as an important end product of scholarship in themselves, in addition to publications. Data are much less “self-archiving” than are publications, however, and often require more human effort to describe and to provide context for interpretation (Borgman, 2007; Cyberinfrastructure Vision for 21st Century Discovery, 2007).

Problem Statement

Researchers want to capture, manage, and store their data in ways that assures its integrity. Similarly, librarians, archivists, and future users of those data want to be assured of the integrity of the data over time. But what does “integrity” mean in this environment, and to whom? We draw upon our research on data practices conducted at a large, collaborative, multi-disciplinary science and technology center to compare the notions of “integrity” held by different research participants to suggest implications for iSchool-based research on information infrastructure and on digital preservation (Wallis et al., 2007).

Research on scientific data practices has concentrated on big science such as physics (Galison, 1997; Traweek, 1992) or on large collaborations in areas such as biodiversity (Bowker, 2000a; b; c). Equally important in understanding scientific data practices is the study of small teams that produce observations of long-term, multi-disciplinary, and international value, such as those in the environmental sciences. The emergence of technology such as wireless sensing systems has contributed to an increase in the volume and complexity of data that can be generated by small research teams. Scientists can perform much more comprehensive spatial and temporal in-situ sensing of environments than is possible with manual field methods, opening up new directions for research questions and methods. Digital library systems not only must be able to integrate disparate data sets that were collected with different research questions in mind, but must also ensure the integrity and usability of those data.

The diversity of research questions, techniques, and instrumentation complicates the long-term preservation and access to data in data digital libraries. Data digital libraries are only as valuable as the data they contain. When data are submitted to prestigious repositories (e.g., Protein Data Bank, 2006), they are evaluated rigorously. This is not the case with data that are made available through local websites or local repositories, where mechanisms for data authentication are less consistent or entirely non-existent. Scientific researchers often prefer to use their own data because they are intimately familiar with how those data were collected, the actions that were taken in the field to collect them, what went wrong and what was done to fix those problems, the context in which the data were collected, and local subtleties and quirks. A prime goal of

data digital libraries is to facilitate widespread use of data by any interested user. What can digital libraries systems do to ensure data integrity? Researchers (or teachers or students) who wish to reuse data rely on a variety of indicators when assessing data collected by others. Indicators include the reputation of the data collector and the institution, the quality of papers reporting the data, instrumentation descriptions and specifications, and any accompanying documentation. When these indicators are not available to users, the task of assessing data integrity is much more difficult. This is especially the case with data in the environmental sciences, where data practices vary widely from project to project. Ecological research questions are typically focused on specific locales or on particular types of phenomena. Because of this, research methods and instrumentation varies widely between projects. Even multi-site comparatives, such as Long-Term Ecological Research projects (LTERs), are not immune from this diversity, as the highly unpredictable and variable nature of in-field ecological research demands flexibility. Researchers must be able to design customized solutions to local and unforeseen challenges (Karasti, Baker & Halkola, 2006).

Research Context

To examine these issues, we are studying the data practices of researchers within the Center for Embedded Networked Sensing (CENS), a National Science Foundation Science and Technology Center established in 2002 [<http://www.cens.ucla.edu/>]. CENS supports multi-disciplinary collaborations among faculty, students, and staff of five partner universities across disciplines ranging from computer science to biology. The Center's goals are to develop and implement wireless sensing systems and to apply this technology to address questions in four scientific areas: habitat ecology, marine microbiology, environmental contaminant transport, and seismology. Application of this technology already has been shown to reveal patterns and phenomena that were not previously observable. CENS' immediate concerns for data management, its commitment to sharing research data, and its interdisciplinary collaborations make it an ideal environment in which to study scientific data practices and to construct digital library architecture to support the use and reuse of research data.

Data integrity issues became more salient within CENS as the technology and the science each matured. In the initial stages of CENS research, the primary focus was developing new wireless sensing system technology. As is typically the case with new technology development, the first few generations of systems were often unreliable, sometimes producing data with scientific value and other times not. CENS has now reached the point where the nascent technology is consistently producing data of real scientific value, and as such must now take means to address long-term data preservation and access. Until recently, CENS relied on a largely oral culture for the exchange of data, and for the exchange of information about how data are collected, such as the equipment used and the state of the equipment. As the Center has grown, an oral culture is no longer sufficient to capture and retain institutional memory. The student research population turns over rapidly and tacit knowledge needs to be exchanged within and between many more research teams.

Future uses of CENS data depend on identifying and implementing solutions for data preservation and access challenges. We have identified a number of critical research

areas regarding CENS data practices. These include research into characterizing field deployment practices, data versions/states, collaboration practices, and requirements for digital libraries based on the needs of the community. Research in these areas is ongoing or in the development stages and will be reported on at the conference.

Brief Summary of Empirical Findings

Digital libraries can facilitate data integrity by recognizing and accounting for the scientific practices and requirements. Scientists have established methods for describing the network, sensors, and calibrations, but often this information is documented separately from the data, if it is documented at all. Among the many research questions provoked by our research are how digital libraries can store essential contextual information and associate it with relevant data points. Sensor faults have a huge impact on the quality and quantity of data generated by wireless sensing system deployments. Similarly, we are concerned with how sensor fault detection can be reflected in digital libraries. Calibration information is essential to post-deployment data analysis, but calibration information varies for each type of sensor, and in some circumstances even between sensors of the same type on the same deployment. Issues arise such as what level of granularity in the calibration information needs to be associated with each data set. Future architecture for wireless sensing systems must address capturing, organizing, and accessing this information.

Implications for iSchool Research

The implications of our research on data integrity are multi-fold for the iSchool research community, and we hope to promote discussion of these issues. iSchool research has focused much more on documents than on data. iSchool expertise in data practices research, and in the preservation and curation of data, is spotty at best. Data are a growing component of the scholarly information infrastructure and must be integrated into larger discussions of technology, institutions, practices, and policy. Integrity of content is a generic issue in curation and preservation. As these concerns come to the fore in data practices, lessons learned here will apply across the scholarly information infrastructure.

References

- Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Bowker, G. C. (2000a). Biodiversity datadiversity. *Social Studies of Science*, 30(5): 643-683.
- Bowker, G. C. (2000b). Mapping biodiversity. *International Journal of Geographical Information Science*, 14(8): 739-754.
- Bowker, G. C. (2000c). Work and information practices in the sciences of biodiversity. VLDB 2000, Proceedings of 26th international conference on very large data bases, Cairo, Egypt. Kaufmann: 693-696.
- Cyberinfrastructure Vision for 21st Century Discovery (2007). National Science Foundation. Visited <http://www.nsf.gov/pubs/2007/nsf0728/> on 17 July 2007.

- Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago: University of Chicago Press.
- Karasti, H., Baker, K. & Halkola, E. (2006). Enriching the notion of data curation in e-Science: Data managing and information infrastructuring in the Long Term Ecological Research (LTER) Network. *Journal of Computer Supported Cooperative Work*, 15(4): 321-358.
- Protein Data Bank (2006). Visited <http://www.rcsb.org/pdb/> on 4 October 2006.
- Traweek, S. (1992). *Beamtimes and lifetimes : the world of high energy physicists*, (1st Harvard University Press pbk. ed.). Cambridge, Mass.: Harvard University Press.
- Wallis, J. C., Borgman, C. L., Mayernik, M., Pepe, A., Ramanathan, N. & Hansen, M. (2007). Know Thy Sensor: Trust, Data Quality, and Data Integrity in Scientific Digital Libraries. 11th European Conference on Digital Libraries, Budapest, Hungary. Berlin: Springer. LINC5 4675: 380-391.

Whose data do you trust? Integrity issues in the preservation of scientific data

Matthew S. Mayernik
Dept of Information Studies
Graduate School of
Education & Information
Studies, UCLA
00+1+3102060029
mattmayernik@ucla.edu

Jillian C. Wallis
Center for
Embedded
Networked Sensing
UCLA
00+1+3102060029
jwallisi@ucla.edu

Alberto Pepe
Dept of Information
Studies
Graduate School of
Education &
Information
Studies, UCLA
00+1+3102060029
apepe@ucla.edu

Christine L. Borgman
Dept of Information Studies
Graduate School of
Education & Information
Studies, UCLA
00+1+3108256164
borgman@gseis.ucla.edu

ABSTRACT

Integrity of content is a generic issue in curation and preservation, but has not been extensively studied in relation to scientific data. Data are now being seen as an important end product of scholarship in themselves. In this paper, we will discuss data integrity issues in relation to environmental and ecological data, and the implications of these issues on the development of data digital libraries. For users to trust and interpret the data in scientific digital libraries, they must be able to assess the integrity of those data. Criteria for data integrity vary by context, by scientific problem, by individual, and a variety of other factors. The goal of this research is to identify functional requirements for digital libraries of scientific data, encompassing both technical and social factors that can affect data integrity. Mechanisms to ensure data integrity have to be present at each stage in the data life cycle, from data collection to data preservation and curation. The implications of our research on data integrity are multi-fold for the iSchool research community, and we hope to promote discussion of these issues.

Topics

Information infrastructure development
Information management
Information technology and services
Nature and scope of iSchools and iResearch
Preserving digital information

Keywords

Data integrity, digital libraries, scientific data practices

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse this work. The authors are solely responsible for their paper's content. Our thanks to the Association for Computing Machinery for permission to adapt and use their template for the iSchools 2008 Conference.

1. INTRODUCTION

Preserving digital information is a central concern to the design of information infrastructure. Digital information takes many forms, and one of increasing importance is scientific data. Data are much less "self-archiving" than are publications, however, and often require more human effort to describe and to provide context for interpretation [5, 10]. When engaged in their own research, scientific researchers take steps to ensure the quality and integrity of the data they capture. However, it is much more difficult for researchers to evaluate data that are collected by others because of the lack of standardization in the ways that data are documented and annotated. These standards are necessary to provide enough information to enable reuse of data by others.

Additionally, researchers will only use data collected by others if they are confident in its quality and integrity. The lack of sufficient methods for ensuring data integrity is a serious impediment to the establishment and use of data repositories in many sciences. In this paper, we will discuss data integrity issues in relation to environmental and ecological data, and the implications of these issues on the development of data digital libraries.

2. PROBLEM STATEMENT

Researchers want to capture, manage, and store their data in ways that assures its integrity. Similarly, librarians, archivists, and future users of those data want to be assured of the integrity of the data over time. But what does "integrity" mean in this environment, and to whom? We draw upon our research on data practices conducted at a large, collaborative, multi-disciplinary science and technology center to compare the notions of "integrity" held by different research participants to suggest implications for iSchool-based research on information infrastructure and on digital preservation [24].

2.1 Scientific Data Practice Research

Research on scientific data practices has concentrated on big science such as physics [13, 22] or on large collaborations in areas such as biodiversity [7, 8, 9]. Equally important in understanding scientific data practices is the study of small teams that produce observations of long-term, multi-disciplinary, and international value, such as those in the environmental sciences.

The emergence of technology such as wireless sensing systems has contributed to an increase in the volume and complexity of data that can be generated by small research teams. Scientists can perform much more comprehensive spatial and temporal in-situ sensing of environments than is possible with manual field methods, opening up new directions for research questions and methods. However, because the use of wireless sensing systems in environmental and ecological research is a relatively new phenomenon, data management techniques for data from such systems are largely local and idiosyncratic in nature.

2.2 Data Standards

Several XML-based standards and protocols exist for this diverse community, but none of them are stable or widely adopted. Structures most relevant to embedded sensor network data in the environmental sciences include the Ecological Metadata Language, supported by the Knowledge Network for Biocomplexity [11, 16], and the Sensor Modeling Language (SensorML), supported by the Open Geospatial Consortium, which describes sensor network equipment and relationships. SensorML is complemented by the Observations and Modeling (O&M) language to express ecology data captured by the sensor network. SensorML and O&M are in the final stages of being accepted as formal standards [23].

The multiplicity of standards in this field poses significant challenges to researchers and has limited the widespread implementation of any individual standard. As the use of complex instrumentation such as wireless sensing systems becomes more widespread, environmental scientists face the questions of what to standardize, when, and for what purposes. Attempts to develop data digital libraries for these domains are hindered by this lack of standardization.

2.3 Data Digital Libraries

Data digital libraries are only as valuable as the data they contain. When data are submitted to prestigious repositories (e.g. [21]) they are evaluated rigorously. This is not the case with data that are made available through local websites or local repositories, where mechanisms for data authentication are less consistent or entirely non-existent. Digital library systems not only must be able to integrate disparate data sets that were collected with different research questions in mind, but must also ensure the integrity and usability of those data.

The diversity of research questions, techniques, and instrumentation complicates the long-term preservation and access to data in data digital libraries. Scientific researchers often prefer to use their own data because they are intimately familiar with how those data were collected, the actions that were taken in the field to collect them, what went wrong and what was done to fix those problems, the context in which the data were collected, and local subtleties and quirks.

A prime goal of data digital libraries is to facilitate widespread use of data by any interested user. What can digital libraries systems do to ensure data integrity? Researchers (or teachers or students) who wish to reuse data rely on a variety of indicators when assessing data collected by others. Indicators include the reputation of the data collector and the institution, the quality of papers reporting the data, instrumentation descriptions and specifications, and any accompanying documentation. When these indicators are not available to users, the task of assessing data integrity is much more difficult.

2.4 Assessing Data Integrity

Assessing data integrity is especially difficult in the environmental sciences, where data practices vary widely from project to project. Ecological research questions are typically focused on specific locales or on particular types of phenomena. Because of this, research methods and instrumentation varies widely between projects. Even multi-site comparatives, such as the Long-Term Ecological Research program (LTERs), are not immune from this diversity, as the highly unpredictable and variable nature of in-field ecological research demands flexibility.

Karasti, Baker, & Halkola describe their work in the LTER program [15]. They have found that it is necessary to give researchers multiple paths and incentives for producing metadata and documentation. Rapid technological change requires the LTER program to take a science-driven view of data management, rather than the technology-driven view that motivates most discussion around e-Science data curation (where the focus is on the "digital obsolescence" problem and digital preservation techniques like migration and emulation) because data regularly outlive the technology used to create them. They emphasize an overall framework for "data stewardship", involving "data definitions, data requirements, and quality assurance as well as user feedback, redesign, and data exchange" (pg. 352).

Researchers in the ecological and environmental sciences must be able to design customized solutions to local and unforeseen challenges. Because of this, assessing the integrity of data is a complex and difficult task.

3. RESEARCH CONTEXT

To examine these issues, we are studying the data practices of researchers within the Center for Embedded Networked Sensing (CENS), a National Science Foundation Science and Technology Center established in 2002 [<http://www.cens.ucla.edu/>]. CENS supports multi-disciplinary collaborations among faculty, students, and staff of five partner universities across disciplines ranging from computer science to biology. The Center's goals are to develop and implement wireless sensing systems and to apply this technology to address questions in a variety of scientific areas, including terrestrial ecology, marine microbiology, environmental contaminant transport, and seismology. Application of this technology already has been shown to reveal patterns and phenomena that were not previously observable. CENS' immediate concerns for data management, its commitment to sharing research data, and its interdisciplinary collaborations make it an ideal environment in which to study scientific data practices and to construct digital library architecture to support the use and reuse of research data.

Data integrity issues are becoming more salient within CENS as the new wireless sensing technologies and the scientific applications using those technologies have each matured. In the initial stages of CENS research, the primary focus was developing new wireless sensing system technology. As is typically the case with new technology development, the first few generations of systems were often unreliable, sometimes producing data with scientific value and other times not. CENS has now reached the point where the nascent technology is consistently producing data of real scientific value, and as such must now take means to address long-term data preservation and access issues. Additionally, questions about the integrity of data are coming to the fore as more useful data are being produced.

Until recently, CENS relied on a largely oral culture for the exchange of data, and for the exchange of information about how data are collected, such as the equipment used and the state of the equipment. As the Center has grown, an oral culture is no longer sufficient to capture and retain institutional memory. The student research population turns over rapidly and tacit knowledge needs to be exchanged within and between many more research teams. Future uses of CENS data depend on identifying and implementing solutions for data preservation and access challenges.

4. BACKGROUND ON DATA INTEGRITY

CENS data result from sensor deployments where researchers deploy sensors and wireless communication equipment in real-world field locations. Assessing the integrity of data captured during real-world sensor deployments is a complex task, encompassing the entire data life cycle: experimental design, equipment calibration, data capture, cleaning, derivation, integration, analysis, publication, and preservation [25]. The following sections illustrate how the task of ensuring data integrity is both a technical and a social process [24].

4.1 Technical Data Integrity Issues

Within CENS, a group of computer scientists, statisticians, and electrical engineering researchers have formed the "CENS integrity group" devoted to looking at the technical aspects of wireless sensing system data integrity. The CENS integrity group focuses on sensor fault detection - that is, detecting when sensor output does not accurately represent the phenomena being measured. Fault detection is often viewed as a step in post-deployment data analysis, where data is rejected during analysis if it is deemed to be faulty. However, this approach is flawed for a couple of reasons. First, it may not be possible to tell what data are faulty after the fact, and second, the amount of data available after a challenging deployment may be so small that none can be spared. This is particularly the case on experiments where sensors are short-lived and require frequent calibration. CENS researchers have had to discard as much as 40% of the data collected on a particular deployment, limiting the amount of scientific analysis possible.

For these reasons, the CENS integrity group is researching fault detection methods that would identify sensor faults as data is being captured in the field. There are many challenges in this task, as sensor faults can take many forms, such as out of range values, and have a number of possible causes, each requiring a different detection method [19]. Often faults compound each other, such as when a failing battery causes a sensor to give faulty readings. This technical data integrity work is important because it indicates that the embedded networked sensing community is becoming aware of the complexity of data integrity issues, even if the mechanisms and techniques they are developing for automated and human-mediated data integrity checks in sensor systems are in the very beginning stages. Once these mechanisms mature, they will be an integral component of scientific research that uses sensor networks.

4.2 Social Data Integrity Issues

Sensor faults represent a technical impediment to ensuring data integrity. Equally important social, cultural, and economic impediments must also be addressed. Bimholtz and Bietz [4] point out that many factors play into data sharing practices. Understanding a data set requires knowledge of the

context in which it was collected. Documenting the data collection process is challenging, however, because much of the knowledge that goes into collecting and interpreting a data set is tacit. Knowledge transfer "is not simply a matter of sharing a set of instructions, but is a highly social process of learning practices that are not easily documented" (pg. 340). They identify three recommendations for the design of systems to share scientific data: 1) Support social interaction around data abstractions and the data themselves, 2) Do not rely on metadata alone, it is also necessary to support the sharing of supplementary materials that enhance the value of the data, and 3) Support social and scientific roles of data. Because tacit knowledge is such an integral part of ecological research, producing data set documentation is time-consuming and labor-intensive process.

4.3 Cultural Data Integrity Issues

Data collections in the environmental and ecological sciences are primarily research data collections [18] that were collected for a specific purpose, and are typically held by the researchers who created them [26]. It is therefore often difficult for researchers to discover and access research data that was collected by someone else. As Zimmerman notes, ecologists face a litany of challenges in finding data. Because few ecological databases exist, the ecologists Zimmerman studied looked to many other sources for data, including peer-reviewed publications, museums, and personal contacts. Once relevant data was found, they went "to great lengths to ensure that they understand data collected by others" [26, pg. 9], relying on their own field expertise to evaluate the potential problems with given data.

Additionally, when combining disparate data sets, researchers often ignored small differences between data sets, with the hope that small amounts of bad data or incompatibilities between data sets will not affect how they can be analyzed. Peer-reviewed literature are generally a good source of information about data, but peer-review typically does not certify data quality per se, as publications rarely report the actual raw data on which they are based. Similarly, databases are challenging to use because they often do not provide enough information to determine the purposes for which data was collected.

Despite the challenges, scientists find ways to get access to the data they need [26]. However, tremendous amounts of time and energy are spent in the process of finding, evaluating, and integrating data. Better methods for data documentation and access would make it much easier for scientists to discover and share data, and, just as importantly, evaluate the integrity of data collected by others. Understanding the requirements for such methods is the focus of our research.

5. RESEARCH METHODOLOGY

We have identified a number of critical research areas regarding CENS data practices, and are pursuing a multi-pronged research approach. We recently completed a series of interviews with scientists from five environmental science projects within CENS. For each project we interviewed a complementary set of science and technology participants, including faculty, post-doctoral fellows, graduate students and research staff. We interviewed 22 participants, each for 45 minutes to two hours; interviews averaged 60 minutes. Interview questions were grouped into these four categories: data characteristics, data sharing, data policy, and data architecture. The interviews were audiotaped, transcribed,

and complemented by the interviewers' memos on topics and themes [17]. Analysis proceeded to identify emergent themes. This study used the methods of grounded theory [14] to identify themes and to test them in the full corpus of interview transcripts and notes.

The next planned stages of our research are based on findings and gaps identified by this interview study, and encompass ethnographic research, information system design, and quantitative characterization of CENS collaborative activities. We are developing another interview study that will focus on data versioning and provenance within CENS research. This study is in the development stages, but will seek to determine in what states or versions data sets exist, and how these states are documented. Additionally, we are examining ways that wireless sensing information and data can be organized through the development of taxonomies, ontologies, or metadata models.

6. DISCUSSION - IMPLICATIONS FOR DATA DIGITAL LIBRARY DESIGN

Digital libraries can facilitate data integrity by recognizing and accounting for the scientific practices and requirements. Scientists have established methods for describing their communication networks, sensors, and equipment calibrations, but often this information is documented separately from the data, if it is documented at all. Among the many research questions provoked by our research are how digital libraries can store essential contextual information and associate it with relevant data points.

6.1 The Context of CENS Data Collection

Within CENS, data collection largely takes place on real-world sensor deployments. We have designed and implemented the CENS Deployment Center (CENSDC), a database for CENS deployment information. The CENSDC provides a centralized web-accessible location for researchers to articulate and document deployment activities through the creation of pre-deployment plans and post-deployment feedback/notes. CENS researchers collaboratively plan sensor deployments prior to going out in the field. Information captured in deployment plans includes deployment dates, locations, participants, technology, equipment, tasks, and other notes. The system also facilitates post-deployment information capture, by providing a series of fields for researchers to outline any problems they encountered while in the field, as well as recommendations and suggestions for future deployments.

The development of the CENSDC has followed an iterative rapid-prototyping design process. Requirements for the system were developed through ethnographic study of CENS deployments and discussions with CENS researchers to characterize deployment practices. Much of the in-field data collection process during deployment involves tacit knowledge about equipment setups, deployment locations, and field preparations from past deployments. As CENS technologies mature and current researchers gain deployment experience, new students face a steeper learning curve when joining a project. The CENSDC was designed to capture tacit knowledge and contextual information surrounding in-field data capture, and to serve as a tool for transfer of common knowledge — knowledge gained through field experiences that can be utilized for future deployments. The CENSDC adds value to CENS data by providing a source of contextual information surrounding the data collection. This information can assist researchers in writing papers, proposals,

and reviews, as well as in maintaining their data and leveraging them for reuse by others.

6.2 Data Integrity and the Data Life Cycle

We have proposed that digital library services should serve scientists whose data exists in all stages of the data life cycle [6, 25]. Building on this model, mechanisms to ensure data integrity will also have to be present at each stage in the data life cycle.

Prior to data capture, equipment are tested and calibrated. Calibration information is essential to post-deployment data analysis, but calibration information varies for each type of sensor, and in some circumstances even between sensors of the same type on the same deployment. Issues arise such as the level of granularity in the calibration information needs to be associated with each data set. At the point of data capture, it will be essential that data digital library implementations accommodate and (ideally) incorporate automated or human-mediated data integrity checks as data are being collected. Sensor faults have a huge impact on the quality and quantity of data generated by wireless sensing system deployments, and researchers must be able to indicate the presence of these faults and their impact on the resulting data.

Similarly, contextual information around the data capture, such as the equipment and software used, is critical to evaluate data. Often this information is not documented on data. Contextual information about the data collection process is particularly important when evaluating data collected by someone else. During the data analysis phase, data sets often undergo changes as scientists clean, integrate, and analyze data. The provenance of these changes is relevant when assessing the integrity of the resulting data, as it may be necessary to backtrack through the data analysis steps that led to a research claim.

Publications are currently the main product of scientific research, but as data and other scientific information is increasingly available online, it would be very valuable to be able to identify relationships between resources, making the scholarly value chain explicit [20]. Finally, curating data and providing effective data stewardship is essential if data are to be used and re-used in the future.

7. CONCLUSION - IMPLICATIONS FOR ISCHOOL RESEARCH

The implications of our research on data integrity 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 [2, 3]. 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" [12], and to facilitate the development of a trained workforce of data scientists [18].

Integrity of content is a generic issue in curation and preservation, but has not been extensively studied in relation to scientific data. As data are increasingly being made available on the internet,

questions about ensuring data integrity across all stages of the data life cycle must be answered. Data that are collected on by ecological and environmental researchers have scientific value both to immediate research questions and long-term longitudinal studies. Distributed longitudinal studies will require standards that ensure the interoperability of the sensor, the network, and the data [1]. These concerns are coming to the fore in data practices; lessons learned here will apply across the scholarly information infrastructure.

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. CENSEI, under which much of this research was conducted, is funded by National Science Foundation grant #ESI-0352572, William A. Sandoval, Principal Investigator and Christine L. Borgman, co-Principal Investigator. Alberto Pepe's participation in this research is supported by a gift from the Microsoft Technical Computing Initiative. Development support for the CENSDC provided by Margo Reviel, Christina Patterson, and Andrew Tai of UCLA Academic Technology Services.

9. REFERENCES

- [1] Arzberger, P. 2004. Sensors for environmental observatories. Report of the NSF sponsored workshop, December 2004, WTEC, Baltimore, MD. <http://www.wtec.org/seo/final>
- [2] Arzberger, P., Schroeder, P., et al. 2004. An International Framework to Promote Access to Data. *Science*, 303: 1777-8.
- [3] Atkins, D., et al. 2003. National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, Revolutionizing Science and Engineering through Cyberinfrastructure. http://www.communitytechnology.org/nsf_ci_report/
- [4] Birnholtz, J. P. and Bietz, M. J. 2003. Data at work: supporting sharing in science and engineering. In M. Tremaine (ed.): GROUP '03. Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work, 2003 November 9 to 12, 2003. ACM Press, pp. 330-348.
- [5] Borgman, C. L. 2007. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- [6] Borgman, C. L., Wallis, J. C., Mayernik, M., and Pepe, A. 2007. Drowning in Data: Digital Library Architecture to Support Scientists' Use of Embedded Sensor Networks. JCDL '07: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, Vancouver, BC. Association for Computing Machinery.
- [7] Bowker, G. C. 2000a. Biodiversity datadiversity. *Social Studies of Science*, 30(5): 643-683.
- [8] Bowker, G. C. 2000b. Mapping biodiversity. *International Journal of Geographical Information Science*, 14(8): 739-754.
- [9] Bowker, G. C. 2000c. Work and information practices in the sciences of biodiversity. VLDB 2000, Proceedings of 26th international conference on very large data bases, Cairo, Egypt. Kaufmann: 693-696.
- [10] Cyberinfrastructure Vision for 21st Century Discovery. 2007. National Science Foundation. <http://www.nsf.gov/pubs/2007/nsf0728/>
- [11] Ecological Metadata Language. 2004. <http://knb.ecoinformatics.org/software/eml/>
- [12] 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.
- [13] Galison, P. 1997. *Image and Logic: A Material Culture of Microphysics*. Chicago: University of Chicago Press.
- [14] Glaser, B. G. and Strauss, A. L. 1967. *The discovery of grounded theory, strategies for qualitative research*. Observations. Chicago: Aldine Pub. Co. x, 271.
- [15] Karasti, H., Baker, K., and Halkola, E. 2006. Enriching the notion of data curation in e-Science: Data managing and information infrastructuring in the Long Term Ecological Research (LTER) Network. *Journal of Computer Supported Cooperative Work*, 15(4): 321-358.
- [16] Knowledge Network for Biocomplexity. 2004. <http://knb.ecoinformatics.org/index.jsp>
- [17] Lofland, J., et al. 2006. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. Belmont, CA: Wadsworth/Thomson Learning.
- [18] Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. 2005. National Science Board. http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf
- [19] Ni, K., et al. in review. Sensor Network Data Fault Types.
- [20] Pepe, A., Borgman, C. L., Wallis, J. C., and Mayernik, M. S. 2007. Knitting a fabric of sensor data resources. ACM/IEEE Information Processing in Sensor Networks Workshop on Data Sharing & Interoperability, Cambridge, MA.
- [21] Protein Data Bank. 2006. <http://www.rcsb.org/pdb/>
- [22] Traweek, S. 1992. *Beamtimes and lifetimes: the world of high energy physicists*, (1st Harvard University Press pbk. ed.). Cambridge, Mass.: Harvard University Press.
- [23] VAST. 2008. Introduction to SensorML. <http://vast.uah.edu/SensorML/>
- [24] Wallis, J. C., Borgman, C. L., Mayernik, M., Pepe, A., Ramanathan, N. and Hansen, M. 2007. Know Thy Sensor: Trust, Data Quality, and Data Integrity in Scientific Digital Libraries. 11th European Conference on Digital Libraries, Budapest, Hungary. Berlin: Springer. LINC5 4675: 380-391.
- [25] Wallis, J. C., Borgman, C. L., Mayernik, M., and Pepe, A. 2007. Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research. 3rd International Digital Curation Conference, Washington D.C.
- [26] Zimmerman, A. 2007. Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *International Journal of Digital Libraries*, 7: 5-16.

FULL PAPER SUBMISSION 11:

Approaches to Semantic Markup for Natural Heritage Literatures

AUTHOR(S):

Cui, Hong

ORGANIZATION(S):

University of Arizona, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Approaches to Semantic Markup for Natural Heritage Literatures (Full Paper Submission)

The theme of this paper is on the application of automated semantic markup techniques on natural heritage literature to address information needs of taxonomists, including the creation, management, and use born-digital or digitized full-text information. Two machine learning based techniques are discussed: one is supervised machine learning and the other is unsupervised.

Since over two hundred years ago, the work has been on-going to document the existence of various living organisms, their habitats, geographic distributions, classifications, and identification keys. As a result of the continuous documentation, a wealth of natural heritage literature has been produced. On-going projects of such include the Flora of North American (FNA) and the Flora of China (FoC). So far the vast majority of natural heritage literature is still in print format. Those produced in electronic format are mainly in html, pdf, or word formats, for example, volumes of FNA and FoC are both in print and html formats. Only a small percentage of such information is in a structured digital format such as a relational database or an XML format (Blum, 2000). To construct the "global commons" of natural heritage literature as envisioned by the Biodiversity Heritage Library (an international collaboration of major natural heritage institutes), a number of digitization projects are being planned or carried out at different levels. Related standard organizations, such as the Biodiversity Information Standards, are working intensively on biodiversity information content/encoding standards, for example the Structured Descriptive Data (SDD, an XML application), to meet the immediate needs of digitization projects. For a review of XML schemas for biodiversity literature, see Sautter et al. (2007).

More and more people from natural heritage communities realize that the transformation of born-digital or digitized biodiversity literature, especially taxonomic descriptions of living organisms, into a machine readable, structured format is an inevitable step to support long term, effective management of the large body of literature. This transformation benefits both of the two major usages of such information: organism identification and information retrieval. For the audience of this conference, structured formats support more accurate retrieval is probably a well-accepted fact. We should also note, however, that descriptions in a good structured format (that allows the encoding of biological characters and their states/values) support faceted browsing as well, in addition to the support for cross collection (i.e. federated) searches. The identification task is traditionally performed by using taxonomic keys, which are decision trees with contrasting statements on a biological character in each node and an identification at the end of a branch. It is worth noting that research (Heidorn & Cui, 2000) has shown that standard full text information retrieval techniques are not effective for specimen identification tasks. The creation of good taxonomic keys for a group of taxons relies on exhaustive search and comparison of descriptions of related organisms and on the completeness of the descriptions. Any taxonomist can testify that neither of the two is supported well by free-text descriptions: it is very easy for an author to leave out some characters and it is also very time-consuming to search for missing characters and compare many free-text descriptions at one time. A structured format can alleviate these difficulties because it allows an information system to remind an author of the missing elements, or auto-fill share characters. It is also possible to make computers shoulder some work involved in comparing descriptions, since once the descriptions are in the structured format, computers can carry out more meaningful comparisons element by element, rather than just matching keywords blindly. If the descriptions are structured in very fine details, such as characters and character states, it even possible to automate many aspects of the generation of taxonomic keys.

A sustainable way for transforming large sets of free text descriptions to structured formats should not rely on grant moneys and graduate assistants' manual labor. Automatic semantic markup techniques must be explored to 1) convert newly digitized full text documents into a structured format 2) allow taxonomists convert a newly created description into the structured format by clicking a button.

The majority of work on structuring taxonomic descriptions focuses on sections of limited semantics, such as taxon names and ranks etc. (Kirkup et. al 2005; Sautter et al. 2006). Far less research focused on cue-poor yet semantic-rich sections, for example morphological description sections, largely due to the lack of consistency in description contents, as shown by Lydon et al.'s (2003). Earlier works (Taylor, 1995; Abascal & Sánchez, 1999; Vanel, 2004) using syntactic parsing methods to extract targeted information to populate relational databases or to mark up plant descriptions in an XML format have been focused on a single collection. Syntactic parsing methods require handcraft grammar rules and extensive lexicons to guide the parsing process. Grammar rules can be collection dependent. Sautter et al. (2007) built an XML editor for biosystematics literature, but it requires the users to come up with regular expression rules to mark up documents. Wood et al. (2004) extracted plant characters from descriptions of five species taken from six floras, using a handmade gazetteer as a lookup list to link extracted terms with their semantic labels. On a much larger scale, Cui and Heidorn (2005) took a supervised machine learning approach to address the issues of inter-collection variations and automatically marked up descriptions of over ten thousands species from three floras down to sentence/clause level. By reusing domain knowledge learned from other collections in the past, their system (called MARTT) achieved better performance on a new collection, without any re-adjustment or fine tuning of the system. The marked-up taxonomic descriptions are currently available for public access at <http://research.sbs.arizona.edu/gs/cgi-bin/library>.

Although the training and mark-up phases of MARTT were automatic and required no human intervention, the training examples came at a cost. MARTT used close to 700 taxonomic descriptions for training. Even though it was just a small fraction (700 out of 17000, or 4%) of total documents processed, it was time consuming and error-prone to manually annotate hundreds of descriptions according to a predefined XML schema (the schema link reveals author's identity, so is not include). On the other hand, the predefined XML schema did not cover all sub-organs described in the descriptions, because a complete schema is only possible after one reads through the entire collections of documents, due to the inherent diversity of living organisms and variations of authorships (e.g. 600+ authors contributed to Flora of China, 800+ to Flora of North America). All uncovered sub-organs were marked as "other-features". Other features include many rarely occurred sub-organs, which are often most valuable information for specimen identification. Treating them indistinguishably as "other features" is not a satisfactory solution. Limited schema coverage is a shortcoming shared by all existing systems reviewed in this article.

The shortcomings of the existing automated markup systems for biodiversity literature prompted us to ask this research question: To what extent the inherent textual cues in the literature themselves can be exploited to produce XML-based markups that are not biased for frequent features.

We started to answer this question by investigating a bootstrapping approach (Riloff and Jones, 1999). This approach is a bottom-up, evidence-based approach, established on the literary warrant principle. It exploits the syntactic cues commonly found in domain corpora. Such cues include: subjects typically start a sentence, subjects are typically nouns or noun phrases, periods are used to end a sentence, while commas or semicolons are used to set off clauses. It is interesting to observe that semantic annotation techniques for so called "general" domains are not applicable for taxonomic descriptions. For example, verbs are the major cues used in information extraction in general domains; however, taxonomic descriptions seldom contain verbs. Phrases such as "such as" and "for example" are major cues used in semantic annotation in general domains for learning equivalent concepts; however, such phrases almost never appear in either FNA or FoC. We had to identify the cues that work for taxonomy domain. Our algorithm automatically learns some concepts using only the strongest evidence and then uses

the newly learned concepts to learn more new concepts. We obtained some promising results on two description collections : one is a collection of 120 algae descriptions, and the other is a collection of 2300+ plant descriptions taken from FNA. From the test collections, the algorithm identifies almost all organs/structures (e.g. in algae description, 1 organ is missed). We also implemented a prototype of a user friendly application that allows taxonomists to mark up their descriptions by a click of a button. The application also allows the users to correct and feedback any error the markup system made. The application is backed up by the knowledge learned by the algorithms. The more knowledge is learned from various collections, the more complete the knowledge will be, and the better performance may be expected. This work is still on-going and we will report the algorithms, the design of the application, and the comparison of the two approaches (supervised vs. unsupervised) at the conference.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

Approaches to Semantic Markup for Natural Heritage Literature

Hong Cui

School of Information Resources and Library Science, University of Arizona.

hongcui@email.arizona.edu

ABSTRACT

The theme of this paper is the application of automated semantic markup techniques on natural heritage literature to address information needs of taxonomists. Two machine learning based techniques (supervised and unsupervised machine learning) are discussed and compared on a real world corpus. A prototype application that supports batch and online modes of converting free text documents to XML format is described.

13.1 Topics

Information organization, Information management, Information technology and services

13.2 Keywords

Semantic markup, natural heritage literature, taxonomic descriptions, biodiversity information, biodiversity informatics, supervised machine learning, unsupervised machine learning, XML

1. Introduction

Since over two hundred years ago, the work has started to systematically document various living organisms, their habitats, geographic distribution, classification, and identification keys. Current projects of such include the Flora of North American [6] (FNA) and the Flora of China [7](FoC). As a result, a wealth of natural heritage literature has been produced. However, the vast majority of natural heritage literature is still in print format. That in electronic format is mainly in html, pdf, or doc formats, for example, volumes of FNA and FoC are both in print and html formats. Only a small percentage of such information is stored in a structured digital format such as relational database or XML format [2]. To construct the “global commons” of natural heritage literature, as envisioned by the Biodiversity Heritage Library (an international collaboration among major natural heritage institutes, <http://www.bhl.si.edu/>), a number of digitization projects are being planned or carried out at different levels. Related standard organizations, such as the Biodiversity Information Standards, are working intensively on a number of biodiversity information content/encoding standards, including the Structured Descriptive Data (SDD, an XML application), to meet the immediate needs of digitization

projects. For a review of XML schemas for biodiversity literature, see [14].

Natural heritage communities have come to the consensus that the reformatting of unstructured biodiversity literature, especially taxonomic descriptions of living organisms, into a machine readable structured format is necessary for the long term and effective management of the large body of literature.

This reformatting benefits both of the two major usages of the information, namely organism identification and information retrieval. For the attendees of the iConference, structured formats enable more accurate retrieval is probably a well-accepted fact that calls for no more elaboration. We should nonetheless note that the organism descriptions in a good structured format (that allows the encoding of biological characters and their states/values) also support faceted browsing and cross collection (i.e. federated) retrieval.

The specimen identification task is traditionally performed by using taxonomic keys, which are decision trees (upside-down trees with the root node on top) with contrasting statements about a biological character in each node and an identification at the end of each branch. It is worth noting that research has shown that full text information retrieval techniques are not effective for specimen identification tasks [8]. The creation of a good taxonomic key for a group of taxa relies on the completeness of the descriptions and exhaustive search and comparison of the descriptions of related organisms. Any taxonomist can testify that neither of the two conditions is supported well by free-text descriptions: it is easy for an author to leave out some characters, it is tedious to repeat shared characteristics again and again in many descriptions, and on top of these, it is also very time-consuming to search for missing characters and compare many free-text descriptions at one time. A structured format can alleviate these difficulties because it allows an information system to remind an author of the missing elements and to auto-fill the shared characteristics. It is also possible for computers to shoulder some work involved in comparing descriptions, thanks to the structures that are made explicit to computers, enabling more meaningful comparisons element by element, rather than

blind keyword matching. If the descriptions are marked-up in very fine details, for example, to the level of characters and character states, it is even possible to automate many aspects of the generation of taxonomic keys.

A sustainable way to transform large sets of free text descriptions to structured formats should not rely on grant moneys and graduate assistants' manual labor. Automatic semantic markup techniques must be explored to 1) convert newly digitized full text documents into a structured format and 2) allow taxonomists to convert a freshly composed description into the structured format by clicking a button.

2. Related Research

The majority of work on structuring taxonomic descriptions focuses on the paragraphs with limited semantics, such as taxon names and ranks etc. [9, 10, 13]. Far less research focused on cue-poor yet semantic-rich sections, for example morphological description sections, largely due to the lack of consistency in description contents, as shown by Lydon et al.[11]. Lydon et al manually compared descriptions of the same five species from six different floras and found large variations in terms of information content and presentation style: only 9% of information was exactly the same, over 55% of information was from a single flora, and around 1% of information from different floras contradicted with each other. Earlier works [16, 1, 17] using syntactic parsing methods to extract information to populate relational databases or to mark up plant descriptions in an XML format have focused on a single collection. Syntactic parsing methods require handcraft grammar rules and extensive lexicons to guide the parsing process. Grammar rules can be collection-dependent, therefore the usefulness of a parser for different collections can be limited. A specialized XML editor for biosystematics literature has also been built [14], however, it requires the users to come up with regular expression rules to mark up documents. [18] extracted plant characters from descriptions of five species taken from six floras, using a hand-made gazetteer as a lookup list to link extracted terms with their semantic labels. It remains to be seen to what extent the manual knowledge engineering scales with larger data sets. On a much larger scale, [4] took a supervised machine learning approach to address the issues of inter-collection variations and automatically marked up descriptions of over ten thousands species from three floras down to sentence/clause level. By reusing domain knowledge learned from other collections in the past, their system (called MARTT) achieved better performance on a new collection, without any re-adjustment or fine tuning of the system. The marked-up taxonomic descriptions are currently available for public access at

<http://research.sbs.arizona.edu/gs/cgi-bin/library>.

3. Two Machine Learning Techniques

The machine learning approach has a number of advantages over other existing approaches. This approach does not require the markup or extraction rules to be provided to the system. Composing the rules by hand is difficult even for

trained knowledge engineers or computer scientists, because one cannot easily tell the overall effect of a rule on a large collection—it may work very well with a small portion of descriptions, but results in a net negative effect if applied on the entire collection. If the goal is to equip domain experts, in this case, taxonomists, with the markup tool, this requirement is simply not reasonable. Taking advantage of ever increasing computational power, the machine learning approach, on the contrary, strives to find global optimum rules automatically.

A properly designed learning system also promotes portability of the technique. It is desirable for a system to be able to process different collections of documents of the same or similar domain, with little or no reengineering. A learning system that learns and makes use of domain knowledge is more portable than a system that learns presentation cues, because it is more likely for the documents to share domain knowledge than the presentation styles, such as font, size, color, or html tags. Since a machine learning system learns markup rules from document collections, the markup rules are always up to date and reflect the characteristics of the current collection. In practice, this works better than relying on hand-crafted rules that may not fit the current task well.

There are two main types of inductive learning techniques: supervised and unsupervised learning techniques.

Supervised learning techniques require training examples, from which the learning system derives knowledge, and against which the learning system verifies its learning. Training examples for the MARTT system were marked-up descriptions manually prepared. The number of training examples required by a system varies, depending on the learning algorithm and the heterogeneity of the data collection. Initially the performance of a learning system increases with the number of training examples. At a certain point (typically before 100% accuracy) it typically reaches a performance plateau. When the plateau is reached, the system is said to be trained, as more training examples will not bring in any performance gain. Unsupervised learning techniques do away with training examples; they derive knowledge from the documents themselves by exploring the various regularities embedded in the documents. For either technique, there has to be some source of knowledge that can be exploited and the learning algorithm has to be able to discover something that is previously unknown to it.

13.3 3.1 The Supervised Learning Algorithm

Aforementioned MARTT system is a supervised learning system involving two phases: the training phase and the mark-up phase. In the training phase, the algorithm learns markup rules/models from training examples, while in the mark-up phase the trained algorithm uses the learned rules to mark up new descriptions. The algorithm is illustrated in List 1. Details can be found in [3], where the algorithm was labeled as SCCP.

13.3.1 The Training Phase

Name: Training

Input: Training Examples (TEs, XML documents with nested elements)

Output: Populated Learning Hierarchy (LH)

Algorithm:

Initialize the empty LH with the root node "description"

Foreach TE in TEs

 Extract elements from TE level by level to save element contents in the corresponding nodes in LH.

 If a node does not exist in the hierarchy of LH, create a new sibling or child node in LH to accommodate the new element at the desired level.

End Foreach

Foreach node in LH

 CreateModel(the element contents saved at the node).

End Foreach

Done

Name: CreateModel

Input: element contents of a node or element

Output: learned model

Algorithm:

N-Grams = Word-level unigrams, 2-grams, and 3-grams extracted from the leading words of element contents.

Foreach N-Gram in N-Grams

 support = the number of instances of the element containing the N-Gram / the total number of instances of the element

 confidence = the occurrence of the N-Gram in the element / the total occurrence of the N-Gram

 Save the rule "N-Gram -> element (confidence, support)"

End Foreach

Done

13.3.2 The Mark-up Phase

Name: Mark-up

Input: LH, to-be-marked document collection C

Output: marked-up document collection C

Algorithm:

Foreach document in C

 MarkupDocument(LH, document)

 Marked-up document = Read the complete, marked-up document off the LH

End Foreach

Done

Name: MarkupDocument

Input: LH, to-be-marked document D

Output: LH with marked up segments of D

Algorithm:

If LH contains one leaf node and one leaf node only

 Mark up D with the name of the leaf node

 Save the marked-up D in the node.

 Done

Else

 Read D into the root node of LH

 The root node marks up segments (sentences/clauses) of D with the names of its child nodes by selecting the rules with the highest confidence and support.

 Save marked-up segments in the node.

 Pass marked segments (S) of D to their corresponding child node (N)

 MarkupDocument(N,S)

13.3.3 List 1: Supervised Learning Algorithm

Note the MarkupDocument function is a recursive function that starts from the root node of LH (a tree with a root on the top and many leaf nodes at the bottom) and ends at a leaf node. At each call, the function works on a successively smaller part of the LH until all branches of the tree is traversed and all leaf nodes are reached.

Although the training and mark-up phases of MARTT were automatic and required no human intervention, the training examples came at a cost. MARTT used close to 700 taxonomic descriptions for training, yet more examples could still mean a better performance as a performance plateau was not reached with that many training examples. Even though it was just a small fraction of total documents processed (700 out of 17000, or 4%), it was time consuming and error-prone to prepare hundreds of descriptions according to the predefined XML schema (<http://publish.uwo.ca/~hcu7/research/xmlschema.xsd>).

On the other hand, the predefined XML schema did not cover all organs described in the descriptions, because a complete schema can be created only after one reads through the entire collections of documents, due to the inherent diversity of living organisms and authorships (e.g. 600+ authors contributed to Flora of China, 800+ to Flora of North America). All uncovered organs were marked "other-features". "Other features" included many rarely occurred organs, which are often the most valuable information for specimen identification. Treating them indistinguishably as "other features" is not a satisfactory solution. Limited schema coverage is a shortcoming shared by all existing systems reviewed in this article.

The shortcomings of the existing automated markup systems for biodiversity literature prompted us to ask this research question: To what extent the inherent textual cues in the literature can be exploited to produce XML-based markups that are not biased for frequent features. In other words, is it possible for an unsupervised learning technique to achieve comparable performance as the supervised technique? To what extent the regularities in the text can replace the knowledge manifested in the training examples?

13.4 3.2 The Unsupervised Learning Algorithm

We started to answer these questions by investigating a bootstrapping approach [12]. This markup approach is a bottom-up approach, established on the literary warrant

principle. It exploits the syntactic cues commonly found in domain corpora. Such cues include: sentences typically start with a subject, the subjects are typically nouns or noun phrases, periods are used to end a sentence, while commas or semicolons are used to set off clauses [5]. These cues were used in the unsupervised learning algorithm that marked up to the level comparable to that of MARTT. List 2 shows the main steps of the algorithm.

Name: Unsupervised Learning Algorithm

Input: To be marked-up documents in text format

Output: Marked-up documents in well-formed XML format

Algorithm:

Use the following simple heuristic rule to find a set of words (Nouns) used as nouns in the documents.

Heuristic rule: A word is used as a noun iff the documents contain its singular and plural forms but no verb forms.

Use English grammar rules to distinguish plural forms from singular forms.

SegmentSet = Segment documents into sentences/clauses (i.e. segments) at the punctuation marks: ,;:

SegmentSet1 = segments starting with a plural noun

SegmentSet2 = segments that do not belong to

SegmentSet1

Foreach segment in SegmentSet1

 Extract the word after the first plural noun and save the word in BoundaryWords

End Foreach

(Nouns, BoundaryWords) = Bootstrap(Nouns, BoundaryWords, SegmentSet2)

Foreach segment in SegmentSet

 If the segment's subject main noun can be identified

 tag = the subject

 Else

 tag = "unknown"

 End If

 If the segment ends with any punctuation mark but a period

 tag = tag + "_block"

 End If

 Use tag to mark up the segment

End Foreach

End

Name: Bootstrap

Input: Nouns, BoundaryWords, SegmentSet2

Output: A bigger set of Nouns, a bigger set of BoundaryWords

Algorithm:

Flag_NewDiscovery = false;

Foreach segment in SegmentsSet2

 Generate a subject pattern based on the up to three leading words. For example, given a segment "Sporangiaster rare.", if "rare" is in BoundaryWords but

Sporangiaster is in neither BoundaryWords nor Nouns, the segment's subject pattern is "?B".

End Foreach

Sort (Patterns) //patterns with fewer ? are ranked higher
Do

 Foreach pattern in Patterns

 If (new nouns or boundary words are discovered)

 /*note: depending on the pattern. The previous example pattern "?B" would result in a new discovery: Sporangiaster is a noun. A pattern such as "?B" would not result in a new discovery, because it could be a "NBB", a "NNB", or a "MNB" (M for modifier)*/

 Flag_NewDiscovery = true

 update Nouns or BoundaryWords.

 End If

 End Foreach

While Flag_NewDiscovery

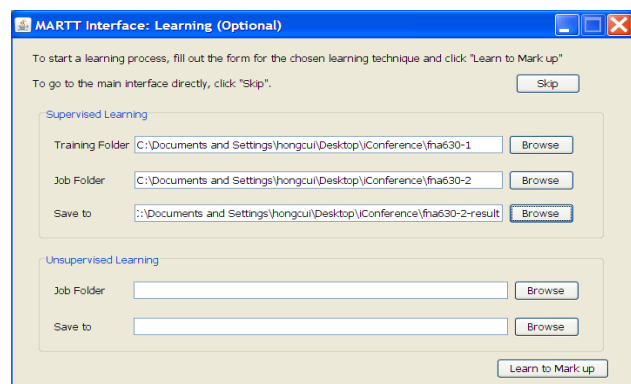
13.4.1 List 2: The Unsupervised Learning Algorithm

The details of the algorithm are still being refined, but the main idea is to start with a small set of Nouns and a small set of BoundaryWords (those are learned from SegmentSet1) to discover more Nouns and BoundaryWords, using the basic assumption "subjects (nouns) are followed by boundary words". The bootstrap function runs in iterations. In each interaction, it tries to make new discoveries and replaces some "?" in the subject patterns with a "B" or "N", until it fails to make any new discovery in an iteration. The learning process stops there. The unsupervised learning algorithm has been tested on a small collection of algae descriptions (120 descriptions), two modest collections of plant descriptions (400-633 descriptions), and a modestly large collection of plant descriptions (2000+ descriptions). The performances on larger collections were better. A later section compares the unsupervised and supervised learning algorithm on the 630 FNA descriptions.

4. The Application

We implemented a prototype of an application that supports both of the learning techniques. The user-friendliness of the application lies not only on the employment of sound interface design principles, but more importantly on the reasonable division of the job between domain experts and the computer. The application does not assume the user has any knowledge engineering skills, such as crafting regular expression patterns for markup, but trusts the user has the ability of identifying misplaced tags. A few screenshots are included here to illustrate the design of the prototype. The application assumes the user either has folders of documents (i.e. batch processing) or will compose a description to mark up. The first screen the user sees once the application is started allows the user to either run a learning algorithm on a folder of documents or to skip the learning step and go directly to the main interface if the

user wishes to access other functions (Figure 1). Note for supervised learning, the user needs to provide training examples. The main interface has functions that help with the preparation of training examples (Figure 3).

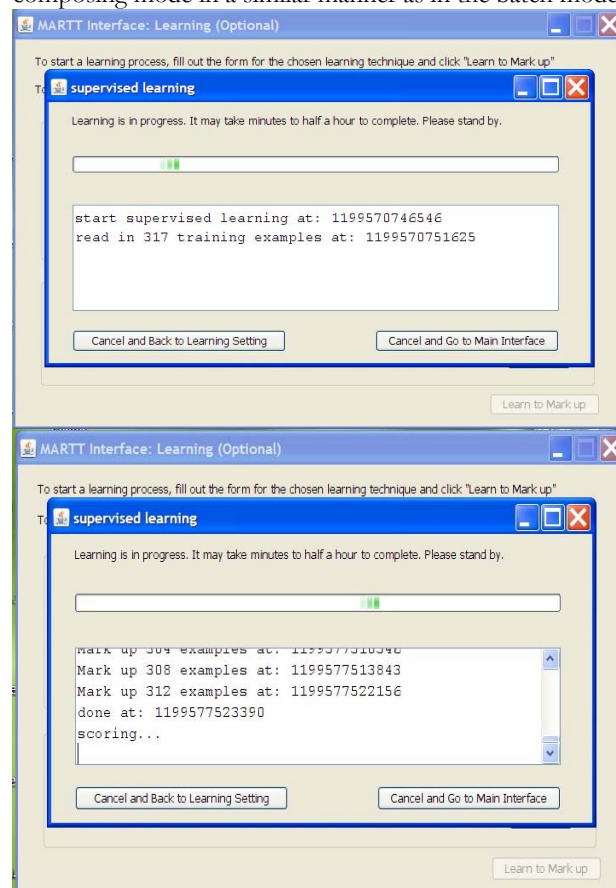


13.4.2 Figure 1: The First Screen Allows

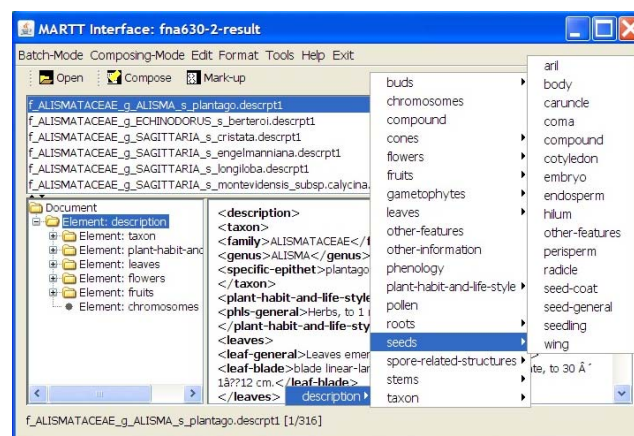
If the user chooses to use either of the algorithms to mark up the collection, she will provide necessary information and instruct the application to “Learn to Mark up”. Since the learning and markup process can take a long time depending on the size of the collection, the interface informs the user the progress (Figure 2). If the documents in the “Job Folder” (i.e. the to-be-marked collection) already contain XML tags, the application will automatically assume the user’s intention is to test the performance of the learning algorithm. It will compare the given tags with those generated by the computer and compute performance scores (Figure 2). The performance scores are saved in a file so the user can compare the performances of different runs. Note once the learning is started, the “Learn to Mark up” button is disabled, but the user is still in control where she would like the application to go. Impatient user may cancel the lengthy learning progress and go either to the learning setting screen (Figure 1) to choose a different learning technique, or to the main interface to use other functions. When the learning and mark up process is done, the user is led to the main interface, where she can verify the correctness of machine-generated markup (using either the supervised or unsupervised approach) by using the “Batch-Mode” functions (Figure 3) and make any corrections by invoking the popup schema menu.

The application maintains the list of files to be examined so the next time the application starts, the user can continue with the remaining of the files. The application automatically records any errors the user corrects, but also allows the user to add any error manually by using “Record an Error” function in “Tools”. A module will be implemented in the future to pass along the identified errors to the learning algorithms so the algorithms can revise its markup rules. In the “Composing-Mode”, the user is presented a text editor window where she can compose a description, which will be marked up in XML format after the user clicks on the “Mark up” button. This one-click markup relies on the knowledge already learned by

MARTT. The user can check for and correct errors in the composing mode in a similar manner as in the batch mode.



13.4.3 Figure 2: Learning Progress



13.4.4 Figure 3: The Main Interface

As mentioned earlier, the interface also supports the preparation of training examples. The user simply opens a list of examples, highlights certain text segments, and tags the segments by selecting appropriate elements from the popup schema menu. Or the user can use the documents

marked-up by the unsupervised algorithm as the training examples for the supervised algorithm.

In short, the application supports the preparation of training examples, two different learning techniques for marking up documents, two different processing modes, the verification and correction of marked up documents, and feedback to the learning algorithms, all in a rather user-friendly manner. An earlier prototype was used by two non-computer science undergraduates and seemed to be straightforward and easy to use. However, being a prototype, there is still room for improvements.

5. Comparison

We ran a simple experiment to demonstrate the differences between the two learning techniques when applied to semantic markup of taxonomic descriptions. We used 633 descriptions from Flora of North America, manually marked-up according to the aforementioned schema. For the supervised learning algorithm, the descriptions were evenly divided into two sets: training and test sets. For the unsupervised learning algorithm, since no training was needed, all descriptions were involved in the bootstrapping. Neither algorithms were optimized for speed at this time, but it took 2.3 times as long for the supervised learning algorithm to mark up 316 descriptions (28.45 mins) as for the unsupervised learning algorithm to mark up 633 descriptions (12.35 mins) on a Gateway 1.20 GHz Internal Core 2 CPU laptop, with the Java heap size set to 512 MB. The supervised learning algorithm put 96% of all words in the 316 descriptions in correct tags. The unsupervised

learning marked 8400 of the 8557 sentences correctly (98%) and identified good tags for 3765 of the 3873 unique subjects. It marked 48 sentences (37 unique subjects) incorrectly and couldn't make a decision on another 109 sentences (71 unique subjects).

Comparing a description for ARISTOLOCHIACEAE *ARISTOLOCHIA serpentaria*, marked up using different techniques (Table 1), we see the effects of the learning algorithms. The unsupervised learning produces well-formed XML documents and the tags are generated and applied as the description contents warrant, while supervised learning assumes a generic XML schema no matter the description contents. The effects of this distinction are shown in different ways: the unsupervised learning generally produced more detailed markup, especially in the leaf description, than the supervised learning; the unsupervised learning explicitly marked organs/structures hidden in "other features" element produced by the supervised learning; although not shown in this example, the unsupervised learning is less constrained in dealing with uncommon organization of descriptive contents, for example, describing multiple organs/structures ("leaves and inflorescence") in one sentence (There were 63 sentences with compound subjects in the 630 FNA descriptions). On the other hand, the XML schema used in supervised learning carries domain knowledge, although implicitly but allowing the translations of concepts at different levels. In this example, it marked up herbs as a type of plant habit and life style, and in

Supervised Markup	Unsupervised Markup
<description>	<description>
<plant-habit-and-life-style> <phls-general>Herbs, erect to decumbent, to 0.6 m.</phls-general> </plant-habit-and-life-style>	<herb>Herbs , erect to decumbent , to 0 . 6 m .</herb> <young_stem>Young stem ridged , glabrous to hispid .</young_stem>
<stems> <stem-general>Young stem ridged, glabrous to hispid.</stem-general> </stems>	<leaf_block> <leaf>Leaves :</leaf> <petiole>petiole 0 . 5 - 3 . 5 cm .</petiole> </leaf_block> <blade_block><leaf_blade>Leaf blade lanceolate to ovate , 5 - 15 * 1 - 5 cm , base truncate to cordate , sinus depth 0 - 1 . 5 cm , apex acute to acuminate ;</leaf_blade> <surface>surfaces abaxially glabrous or hispid ;</surface> <venation>venation pinnate .</venation> </blade_block>
<leaves> <leaf-general>Leaves: petiole 0.5-3.5 cm.</leaf-general> <leaf-blade>Leaf blade lanceolate to ovate, 5-15 Å? 1-5 cm, base truncate to cordate, sinus depth 0-1.5 cm, apex acute to acuminate; surfaces abaxially glabrous or hispid; venation pinnate.</leaf-blade> </leaves>	<inflorescence_block> <inflorescence>Inflorescences from base of stem , an additional flower in axil of stem leaf , racemes ;</inflorescence> <peduncle>peduncle bracteolate , to 1 . 5 cm ;</peduncle> <bracteole>bracteoles lanceolate , to 3 mm .</bracteole> </inflorescence_block>
<flowers> <inflorescence-general>Inflorescences from base of stem, an additional flower in axil of stem leaf, racemes;</inflorescence-general> <peduncle>peduncle bracteolate, to 1.5 cm;</peduncle> <bracteole>bracteoles lanceolate, to 3 mm.</bracteole>	<flower_block> <flower>Flowers :</flower> <calyx>calyx brown_purple , bent ;</calyx> <utricle>utricle pendent , pear_shaped to ovoid , 0 . 5 - 5 cm ;</utricle> <syrinx>syrinx present , ringlike , 1 mm , oblique ;</syrinx> <tube>tube bent , cylindric , 1 cm ;</tube> <annulu>annulus smooth ;</annulu> <limb>limb purplish brown , 3 - lobed , lobes 0 . 5 * 0 . 5 cm , glabrous ;</limb> <gynostemium>gynostemium 3 - lobed , globose to crown_shaped , 1 . 5 mm ;</gynostemium> <anther>anthers 6 ;</anther> <ovary>ovary 3 - locular , to 1 . 5 cm .</ovary> </flower_block>
<flower-general>Flowers: calyx brown-purple, bent;</flower-general> <other-features>utricle pendent, pear-shaped to ovoid, 0.5-5 cm; syrx present, ringlike, 1 mm, oblique; tube bent, cylindric, 1 cm; annulus smooth; limb purplish brown, 3-lobed, lobes 0.5 Å? 0.5 cm, glabrous; gynostemium 3-lobed, globose to crown-shaped, 1.5 mm; </other-features> <anther>anthers 6;</anther> <ovary>ovary 3-locular, to 1.5 cm.</ovary> </flowers>	<capsule_block> <capsule>Capsule globose , 0 . 8 - 2 * 1 - 2 cm , dehiscence basipetal ;</capsule> <valve>valves 6 ;</valve> <septum>septa absent .</septum> </capsule_block>
<fruits> <fruit-general>Capsule globose, 0.8-2 Å? 1-2 cm, dehiscence basipetal;</fruit-general> <valve>valves 6;</valve> <septum>septa absent.</septum> </fruits>	<seed>Seeds rounded , ovate , 0 . 5 * 0 . 4 cm .</seed>
<seeds> <seed-general>Seeds rounded, ovate, 0.5 Å? 0.4 cm.</seed-general> </seeds>	
<chromosomes>2n = 28.</chromosomes>	<2n>2n = 28 .</2n>
</description>	</description>

13.4.5 Table 1: The Comparison of the Same Description Marked-up by Different Learning Algorithms

effect translated <herb> to its broader concept <plant-habit-and-life-style>. Similarly it took <capsule> as a type of <fruit>, and <2n> <chromosome> counts.

There were also 50 instances of organ/structure descriptions marked as “other features” by the supervised algorithm, including those about staminodial ring, utricle, 1st leaves, glandular hairs, intravaginal squamules, pigment cells, stomates, ring meristem, ostiole, air chambers, leaf primordia, terminal leaflet, abaxial ridges, stranded plants, hastula, cells, sporangiasters, staminate scales, and cystoliths, totally 19 organs/structures. Unsupervised learning, on the other hand, discovered good tags for 14 of these structures and marked 40 of these instances with correct tags.

6. Discussions

What have we learned from the experiment on the randomly selected 633 FNA descriptions? First we learned that the documents from the real world will contain information that is not covered by a predefined XML schema. In the 633 descriptions (a tiny portion of FNA, new volumes of which are still being published), 19 organs/structures were not found in the XML schema. In addition, there are also 63 descriptive sentences with compound subjects, which cannot be tagged as any single organ/structure. We believe a semantic schema, let it be an XML schema or an ontology, is necessary for any semantic markup task to ensure consistent interpretation. The top-down approach to the creation of such a schema, however, does not seem to be a viable approach. In order to discover all organs/structures described in a collection, we believe a bottom-up approach based on the literature warrant principle such as the unsupervised learning method described in this paper is more feasible. To say the least, the unsupervised learning method can be used to generate element candidates and to identify difficult cases (“unknown”) for the human experts to base their work on.

Second we saw that the strength and weakness of the two techniques are quite evident. The unsupervised technique is more flexible, more efficient, and capable of uncover new/rare organs/structures. But it has the shortcoming of unable to associate concepts with their broader concepts. For example, it marks different fruit types explicitly yet unable to tell that they share one broader concept “fruit”. We often assume that the semantic relationships among domain concepts have long been codified in a thesaurus or something similar. While such assumptions are often unfounded, even if they are, we still face the issue of discrepancies among different thesauri: for example, if you trace the hypernymy relation of the word nut in WordNet, you will find nut is a seed. However, in the Oxford Plant Characters thesaurus, nut's parent term is fruit. In WordNet, a capsule is a seed vessel, but in Oxford it is a fruit. Fortunately, the unsupervised learning algorithm reduced the workload for human experts to make situated associations among concepts from reading X sentences to examine Y distinct tags. In case of the 633 FNA descriptions, X=8557, Y=280.

This identifies an additional component to be implemented as a part of the MARTT Interface application. In addition to support supervised and unsupervised learning, a module is needed to allow a human expert to establish semantic relationships among existing or introduced concepts. This is what we will do next.

14. 7. References

- [1]. Abascal, R. & Sánchez. (1999) X-tract: Structure Extraction from Botanical Textual Descriptions. Proceeding of the String Processing & Information Retrieval Symposium and International Workshop on Groupware, SPIRE/CRIWG, pp2-7.
- [2]. Blum, S. D. (2000). An Overview of Biodiversity Informatics. Accessed 1/7/2008. http://www.calacademy.org/research/informatics/sblum/pub/biodiv_informatics.html
- [3]. Cui, H. (2005a). MARTT: Using knowledge based approach to automatically mark up plant taxonomic descriptions with XML. Proceedings of the Annual Meeting of American Association of Information and Technology. Oct 28-Nov 2. 2005 Charlotte, North Carolina, USA.
- [4]. Cui, H., & Heidorn, P.B. (2007). 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.
- [5]. FNA Editorial Committee (2006). *Flora of North America North of Mexico Guide for Contributors*. Accessed July 10, 2007 from http://www.fna.org/FNA/Guide/guide_2006.pdf.
- [6]. FNA Flora of North America Editorial Committee. (Eds.). (1993-). *Flora of North America North of Mexico*. Accessed July 10, 2007 from <http://www.fna.org/>
- [7]. FOC Flora of China Editorial Committee. (Eds.). (1994-). *Flora of China*. Beijing/St. Louis: Science Press/Missouri Botanical Garden Press. Accessed July 10, 2007 from <http://flora.huh.harvard.edu/china/>
- [8]. Heidorn, P.B. & Cui, H. (2000). The Interaction of Result Set Display Dimensionality and Cognitive Factors in Information Retrieval Systems. *Proceedings of the Annual Meeting of the American Society for Information Science, ASIS 2002*, pp. 258-270.
- [9]. Kirkup, D., Malcolm, P., Christian, G., & Paton, A. (2005). Towards a digital African Flora. *Taxon*, 54(2). 457-466.
- [10]. Koning, D., Sarkar, I.N., & Moritz, T (2005). TaxonGrad: Extracting Taxonomic Names from Text. *Biodiversity Informatics*. 2. 79-82.
- [11]. Lydon, S.J., Wood, M. M., Huxley, R., & Sutton, D.(2003). Data Patterns in Multiple Botanical Descriptions: implications for automatic processing of legacy data. *Systematics and Biodiversity* 1(2). 151-157.

- [12]. Riloff, E. & Jones, R. (1999). Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping. Proceedings of the 16th National Conference on Artificial Intelligence. pp. 474-479
- [13]. Sautter, G., Agosti, D., & Böhm, K. (2006). A Combining Approach to Find All Taxon Names(FAT). Biodiversity Informatics. 3, 46-58.
- [14]. Sautter, G., Agosti, D., & Böhm, K. (2007). Semi-Automated XML Markup of Biosystematics Legacy Literature with the GoldenGATE Editor, Proceedings of PSB 2007, Wailea, HI, USA. Accessed July 10, 2007 from <http://psb.stanford.edu/psb-online/proceedings/psb07/sautter.pdf>
- [15]. Sautter, G., Böhm, K., & Agosti, D. (2007). A Quantitative Comparison of XML Schemas for Taxonomic Publications. Biodiversity Informatics. 4, 1-13.
- [16]. Taylor, A.(1995). Extracting Knowledge from Biological Descriptions. Proceedings of 2nd International Conference on Building and Sharing Very Large-Scale Knowledge Bases. pp. 114-119.
- [17]. Vanel, J.-M. (2004). Worldwide Botanical Knowledge Base. Accessed July 5, 2007 from <http://wwbota.free.fr/>.
- [18]. Wood, M., Lydon, S., Tablan, V., Maynard, D. & Cunningham, H. (2004). Populating a database from parallel texts using ontology-based information extraction. In Meiziane, F. and M_étais, E., (editors) Proceedings of Natural Language Processing and Information Systems, 9th International Conference on Applications of Natural Languages to Information Systems. pp.254-264.

FULL PAPER SUBMISSION 12:

An Exploration of the Use of Games in Virtual Worlds for Online Education

AUTHOR(S):

Garcia-Murillo, Martha
MacInnes, Ian

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Advances in computing and telecommunications make it possible to take advantage of immersive electronic environments to deliver content. In this paper we present a policy game to be used in a virtual world. The benefits of this tool are examined using Gee's learning principles. From this analysis we find that games in virtual worlds enable reflective exploration that helps participants to learn from their mistakes. Learning takes place from the content conveyed through the game and through the multimedia immersion that allows students to learn the nuances of these virtual contexts. Because there are no real world consequences, participants can take risks, provide or receive help from other students and, most importantly, apply this knowledge to a real world situation. Recommendations are provided to educators to help them exploit the great potential of games while being prepared for the obstacles that they will face.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

An Exploration of the Use of Games in Virtual Worlds for Online Education

Authors: Martha García-Murillo and Ian MacInnes²

Abstract

Advances in computing and telecommunications make it possible to take advantage of immersive electronic environments to deliver content. In this paper we present a policy game to be used in a virtual world. The benefits of this tool are examined using Gee's learning principles. From this analysis we find that games in virtual worlds enable reflective exploration that helps participants to learn from their mistakes. Learning takes place from the content conveyed through the game and through the multimedia immersion that allows students to learn the nuances of these virtual contexts. Because there are no real world consequences, participants can take risks, provide or receive help from other students and, most importantly, apply this

² The authors acknowledge the support of Joe Rubleske, who helped in the selection of the virtual world and designed the roles and the room that for the policy game.

knowledge to a real world situation. Recommendations are provided to educators to help them exploit the great potential of games while being prepared for the obstacles that they will face.

Introduction

The purpose of this paper is to explore whether virtual worlds can provide a setting for a rewarding learning experience for college students. While online college education is a fairly common practice today, most distance interactions with students rely on relatively mature technologies such as course management systems (e.g., WebCT; Blackboard), message boards, electronic mail and weblogs ('blogs'). These applications have facilitated the asynchronous interaction of individuals located in various places and time zones. However, information and communication technologies have advanced considerably, and college professors now have the opportunity to experiment with more innovative software applications that could make online interactions more engaging and stimulating. Virtual worlds represent one such application. In this paper we explore the pedagogical benefits of virtual worlds, which we describe here as graphically immersive, persistent, shared and typically avatar-based digital environments. We believe at the outset that virtual worlds offer instructors a potentially powerful tool for student learning and interaction through simulated experience.

The paper is divided in five main sections. The first section presents evidence from academic research of the potential benefits that virtual worlds can offer to distance students. Here we focus on video games and virtual worlds and the educational benefits that they can provide. The section also identifies the differences between virtual worlds and video games. The second section describes a lobbying game that was adapted from a traditional classroom to a virtual world environment. This was done because of the increasing popularity of online classes at university campuses. There is thus a need to find online activities that maintain the interest of students. Here we describe the simulation and the process that we followed to select a virtual world that could work with this simulation. The third section of the paper analyzes the educational value of the lobbying game within the virtual world. To do this we used Gee's learning principles and determined if virtual worlds offer those benefits. The fourth section presents some of the challenges to educators of the process of developing games for virtual worlds. The fifth and last major section prior to the conclusions talks about future trends of virtual worlds in educational settings.

Video and Computer Games as Pedagogical Tools

Work, leisure and education have all been affected by advances in information and communication technologies. Technologies for entertainment purposes can be so captivating now that traditional classroom and electronic education programs sometimes pale in comparison. The video and computer game industry, in particular, has evolved radically over the past decade to offer interactive capabilities that were only imagined twenty years ago. The cutting-edge animation, opportunities for interaction, and dynamically generated narratives that can be found in today's games have attracted large, diverse audiences, and many American teenagers and young adults play these computer games on a routine basis (Jayakanthan, 2002).

It is thus not surprising that aspects of our lives that were not initially affected by computer-based multimedia are now being transformed. Two such aspects that are being integrated and made increasingly less distinct are entertainment- and learning-related activities. Educational video games, for example, are now being designed around movie titles (Jayakanthan, 2002) and routinely played on home computers (Kerawalla & Crook, 2002). In contrast, computers at many schools are still being used heavily for word processing – an activity that students report finding less than stimulating (Mumtaz, 2001). Given the many mediated outlets that students now have, each one competing for the student's attention, there is greater pressure on educators to be more creative in the delivery of instructional material.

The introduction of computer games in classrooms has aroused the interest of researchers across many fields, and as a result more papers are being published on the subject. Franklin, Peat and Lewis (2003) found that many computer games allow students to realize and react to the consequences of their (and their peers') decisions. Subramanian, Khang and Sai (1999) found that computer games can diminish the apathy and boredom that affect many students. Heffler (2001) and De Vita (2001) argue that computer games can help instructors address differences in learning styles. According to Fleming and Rickwood (2001), video and computer games improve the moods of those who play them. His experiment showed that children who played violent video games experienced increased states of arousal more frequently than those who played pencil-and-paper games, but that their tendency towards violence did not increase. Fleming's study suggests that video and computer games have the potential to lead to more satisfying learning experiences.

These and other results should not be surprising given that video and computer games are widely perceived as vehicles for fun and play, and in such context an association with learning has the potential to increase a person's understanding and retention of the material being taught. With their rich and immersive graphics, interactive capabilities, and users represented visually as avatars (i.e., graphic representations of users in virtual space), virtual worlds are similar to video and computer games in many ways, with a key difference being that the latter impose grand narratives and superseding goals while the former enable users

to pursue their own goals and construct their own narratives. The constructivist paradigm that underlies virtual worlds makes them potentially more enriching for education experiences than traditional video games.

In the process of developing this chapter we needed to keep in mind that the focus of this book is on educational games situated in virtual worlds. In this context we need to think about formal education as a life long process. With the emergence of online education, more and more adults are returning to school to obtain higher degrees, to change careers, or simply to upgrade their skills. These are non-traditional students that already have work experience and their jobs impede them from participating in traditional classrooms. They are instead embracing the flexibility of online education. At the same time many of the professionals who have recently entered the work force are well versed with technologies and, as a result, educators have the opportunity to push the envelop and provide online experiences that challenge them and provide them with unique learning opportunities.

Virtual Worlds

Play is an activity that enhances children's abilities by promoting exploration and experimentation. Children create imagined worlds in their play. As we grow older, though, our opportunities to explore and create imagined worlds are increasingly limited, and our creativity is curbed as a result (Harter, 1981). Virtual worlds offer an alternative space where educators can provide students the opportunity to explore and create. Because they may be perceived as entertainment, virtual worlds may be appealing to distance students as learning environments.

Virtual worlds provide great potential for every level of education, including post-secondary. There are countless examples of educational institutions that have already identified the potential of this means of learning. Ohio University (<http://www.youtube.com/watch?v=aFuNFRie8wA>), Harvard Law School (<http://blogs.law.harvard.edu/cyberone/>) and Case Western Reserve University (<http://blogs.law.harvard.edu/cyberone/>) for example all have campuses in Second Life (<http://secondlife.com/>). In these campuses classes are enhanced with multiple media to enrich the experiences of the online student population. These environments can allow off campus students to feel a closer connection to the university. Teachers and peers can enable experiences that may be difficult to replicate in the real world. For example, a psychiatry professor at the University of California Davis created a room where his students were able to see what patients with schizophrenia often see in their hallucinations (James, 2006).

Because of the versatility of virtual worlds educators need to realize that learning activities will require the development of games. This thus means that much preparation has to occur before the game can be ready for online classroom use. According to (Bartle, 1990), there are several elements that make a game. First it should have a set of rules. If the game entails competition then there should be a way of determining who wins. This can be done by specifying criteria or by identifying objectives to be met at the end. Because the purpose of an educational game is to teach a concept or a skill, the games should allow the participants to learn or practice whatever skills or knowledge is meant to be acquired.

Virtual worlds are often perceived to have a steep learning curve and this is a reason why relatively few educators have made use of virtual worlds. There are some notable exceptions, however. Kusunoki, Sugimoto and Hashizume (Kusunoki, Sugimoto, & Hashizume, 2000) have used a virtual world to teach urban planning and environmental concepts. At the Cornell Theory Center, researchers have developed a virtual world around the idea of a museum as a mechanism to teach science, languages and social studies in a more informal environment (Maher & Corbit, 2002). In the field of health care, researchers have used virtual environments to teach medical concepts (Mantovani, Castelnovo, Gaggioli, & Riva, 2003).

In this paper we will use, Gee's (Gee, 2003) 'learning principles' to show how virtual worlds can enhance learning. We will pair these learning principles to the virtual world to determine if there is pedagogical value in the use of these synthetic environments.

Sample Politics Game for a Virtual World

This paper is a synchronous role-playing simulation based on a game developed by Geoff Wong (Laver, 1979) [see appendix for instructions]. In this game in which each player assumes one of 12 roles (e.g., U.S. Senator, AT&T Vice President, Director of the Business Software Alliance) and tries to convince other players to vote in accordance with his or her role's interests on one or more of three legislative bills. Players are instructed to employ any strategy at their disposal to accomplish this and hopefully those that are outline in the policy section of this paper. The use of actual bills makes it easier for instructors when students ask for more information about them. Our simulation made use of three bills that have recently been considered by one or more legislative bodies:

1. SPY ACT (Securely Protect Yourself Against Cyber Trespass Act). If passed, SPY ACT would impose fines of up to \$3 million against makers of software that steal personal information from a user's computer or hijack a user's browser.
2. Digital Media Consumers' Rights (DMCR) Bill. If passed, the DMCR Bill would (a) allow consumers to break copy controls to make personal copies of audio tracks or movies from discs (CDs, DVDs) they own, and (b) require all copy-protected CDs and DVDs to be labeled as such.

3. Broadband Regulation and Modernization (BRM) Bill. If passed, the BRM Bill would prohibit a government or any entity it creates from offering broadband service for free.

Ideally, the students will begin the simulation knowing a great deal more about these bills than what is provided by these summaries. Two factors work against this ideal, though. First, instructions for playing are relatively long, even with the summaries, and would be made even longer with, say, an appendix detailing the bills' specifications. Players of any game tend to find a thick stack of instructions daunting. Detail must always be balanced with "playability" and a reasonably shallow learning curve. Second, the chief pedagogical aim of the simulation is to teach lobbying concepts. If, in the process, students learn more about the bills used to demonstrate these concepts – and they most certainly will – then the simulation can be considered a success on multiple levels. However, depending on the course being taught, instructors may not have the time to teach students lobbying concepts and the history and particularities of the bills.

One of the most challenging aspects of designing this simulation is the need to establish a "level playing field" at the outset. In other words, at the start of each simulation, each bill should have, to the extent possible, an equal number of roles with explicit voting preferences supporting and opposing it. To run a simulation in which, say, eight players initially oppose a bill serves to disadvantage (and possibly discourage) the four players who favor it.

We created 12 roles that, to varying degrees and in various ways, had a stake in one, two or all three of the bills listed above. We chose to create 12 roles because we believed that it would be difficult to coordinate the conversations and lobbying efforts of a larger group during the simulation. These roles included:

1. The Governor of the state of New York;
2. The Vice President of AOL/Time Warner;
3. The Director of the Computer & Communications Industry Association (CCIA);
4. The Vice President of AT&T;
5. The Director of the New York Association of Cities and Towns (NYACT)³;
6. The Director of the Business Software Alliance (BSA);
7. The Director of the Electronic Frontier Foundation (EFF);
8. The Chancellor of the State University of New York (SUNY);
9. The Director of the American Marketing Association (AMA);
10. The Director of the American Library Association (ALA);
11. The Deputy Secretary of the U.S. Department of Higher Education; and
12. A United States Senator from the state of New York.

A final and important design choice for this simulation is the decision to include or exclude explicit voting preferences for each role at the outset. In other words, instructors can opt to tell students how they (in their roles) intend, at least initially, to vote on each bill, or they can withhold this information and let the students decide this for themselves based on a brief statement of their role's mission, which is provided to them by the instructor. As with the level of detail on selected bills, this decision may be influenced by time constraints, pedagogical objectives, and assumptions about students' abilities and diligence.

One of our chief concerns in this game is that students may not take the time to conduct rudimentary research on their roles and, as a result, could ask us for additional information. Further, we speculated that knowing how one's role initially stands on a bill might help students assume their roles more faithfully and, as a result, make them less inclined to play and vote according to personal preferences towards the bill.

Finally, if one opts, to tell students how their role initially stands on each bill, then it is essential to design the simulation such that each role has, at the outset, an explicitly undecided voting preference, or is said to be merely leaning in one direction, on at least one of the bills. It is hoped that this gives each player greater freedom to try to influence their fellow participants. In this simulation two of the roles– the Governor of New York and the U.S. Senator from New York – are designed as undecided on all three bills, perhaps causing these two roles to wield a disproportionately greater influence on voting outcomes.

Selecting a Virtual World

When working on the game we were inspired to design and conduct an avatar-based virtual simulation to offer students a stimulating alternative to the more conventional activities that are enacted in physical classrooms. We also wanted to break away from the traditional bulletin board discussions of online education. While these conventional modes have been shown to be pedagogically useful, we speculated that a simulation performed by avatars in a graphically rich online environment might possess a distinctive quality that is particularly appealing to undergraduate and graduate students. One of the authors had already designed

³ We invented a body that we named the New York Association of Cities and Towns in order to have a role that represents local interests.

and conducted a simulation for use in a physical classroom, so our tasks involved identifying a suitable virtual world and adapting the simulation for it.

We examined two virtual worlds: Second Life and Habbo Hotel. For our purposes, the power to create and modify an avatar and three-dimensional virtual environment using cutting-edge and relatively sophisticated design tools, though impressive, entails a steep learning curve for beginners. This concern, coupled with the required US\$10 per person membership cost, which has since been waived, led us to consider Sulake Corporation's Habbo Hotel, a virtual world designed for and marketed to teens. Sulake's aim of profitably supporting affirmative online communities for paying teens, though beneficent, was not a factor in our decision to use Habbo Hotel. Instead, we wanted a virtual world that:

1. Lets typical (non-expert) users easily create and use an expressive and visually appealing avatar;
2. Makes it easy for users to communicate through and move their avatars;
3. Provides private virtual space (such as a virtual room) in which members of a small group can interact comfortably and without interruption;
4. Possesses a vibrant interface (or "look and feel") that might appeal to undergraduate and graduate students; and
5. Imposes no monetary costs on the students participating in the simulation.

Sulake's revenues come chiefly from users who purchase Habbo credits to decorate the empty virtual rooms. As Figure 1 illustrates, we furnished the "lobbying room" with assorted rugs and chairs, a blue pinstriped floor, a potted palm tree and two whimsical plaques for the walls, all for US\$20.



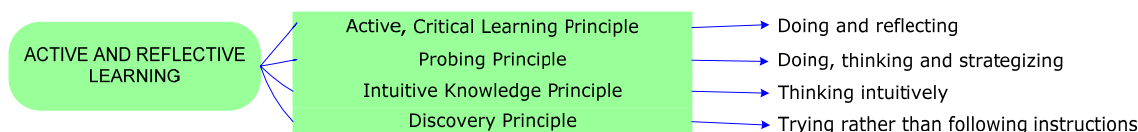
Figure 1: A screen capture of the "Logrolling Room" in Habbo Hotel. An author's avatar sits pensively in one of the chairs.

Habbo Hotel satisfied our other conditions as well. All avatars look like cartoonish depictions of adults but some distinctiveness can be achieved through one's selection of skin color, hair color and style, facial expression, and the style and color of one's shirt, pants/skirt/shorts and shoes. Finally, the authors, each with limited experience in virtual worlds, found moving their avatars and navigating Habbo Hotel relatively simple and straightforward. Thus, despite being somewhat concerned about the reactions of college students to a virtual environment with a somewhat adolescent motif, the authors chose to use a specially designated room in Habbo Hotel as the space in which to conduct the simulation.

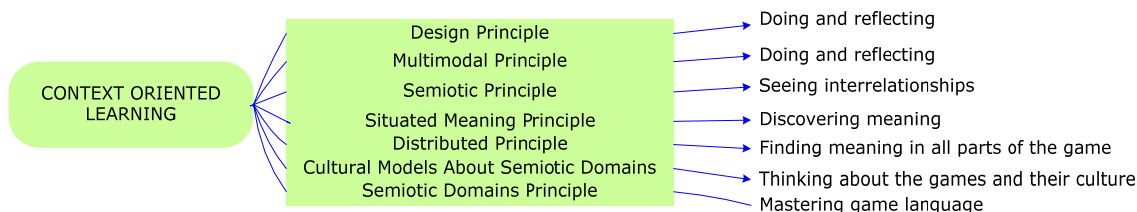
Educational Benefits of Games in Virtual Worlds

While it is clear that students have a lot more sources for entertainment and information this does not necessarily imply that games in virtual worlds would be beneficial for learning. To determine the effectiveness of games in these online environments this section presents an analysis of the learning principles that Gee (Gee, 2003) proposed in his book entitled *What Video Games Have to Teach us About Learning and Literacy*. Many of the 36 learning principles that Gee presents are, however, closely related and to simplify the presentation of the analysis with respect to games and virtual worlds the authors grouped these 36 principles into 8 larger categories.

When Gee wrote his book he was focusing primarily on video games and not necessarily on virtual worlds. The goal of this analysis is to determine if the learning principles that he proposed in the context of video games can also apply to virtual worlds. It is also important to realize that due to the fact that virtual worlds are not games per se, much of the learning that takes place comes from the games or simulations that the instructor prepares in advance for the students. In this paper we described an information policy simulation. Thus the analysis of Gee's principles for this section is done here within the context of that simulation.

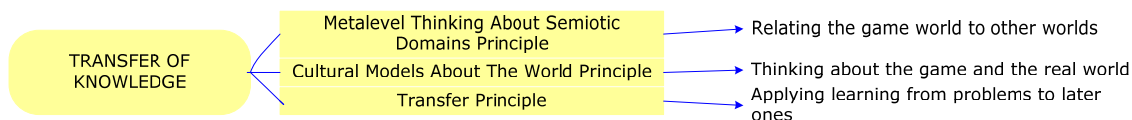


Active and Reflective Learning. The simulation was intended to teach students about the strategies that lobbyists of government officials commonly use to influence policy. In this case each student played the role of a different character and was asked to influence to vote in favor of their character's interests. The simulation used the Habbo Hotel platform. The simulation required the students to identify the viewpoints of other characters and then try to influence them through conversation. The students have the opportunity to experiment with multiple strategies as they participate in the game. They have to discover the positions of those who do not share their views and then think critically about the arguments that they can use to convert them to their side. They can also experiment and strategize to make their influencing efforts more effective. This process of discovery and being actively engaged in the activity help them learn the complexities of the lobbying process first hand.

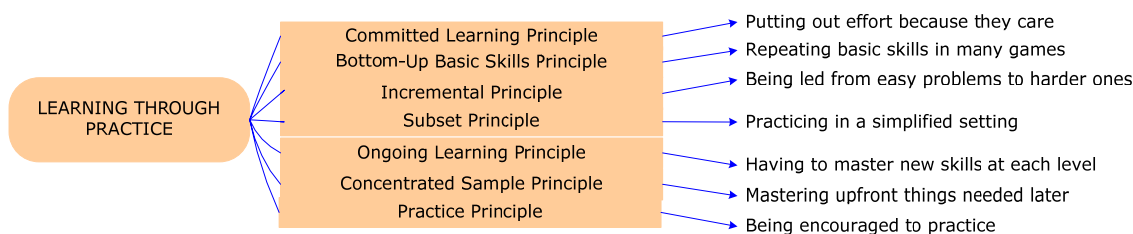


Context Oriented Learning. One of the major advantages of virtual worlds and games is the many opportunities that students have to learn from new environments and be able to adapt to new circumstances. A virtual world offers multiple opportunities for that. First every time that a person enters a virtual world they need to learn from the context itself, the design, other characters, and virtual artifacts. They also have the opportunity of learning from a variety of media such as graphics, text, sounds and even voice. From the context and the different media they need to find the meaning of these elements and the culture of this virtual world to determine norms of conduct. It is through careful observation that students learn these subtleties. In the context of the policy simulation the students need to determine whether it is better to shout, to speak, or to whisper a given comment. Some of these text speaking modes will be acceptable under certain circumstances while others will not. They need to determine if it is better to lobby a group of people at once or somebody alone. They need to determine how to initiate a conversation or how to join a conversation that has already started. All of these norms of conduct are each new opportunities for experimentation and learning. They can become more attuned to small subtleties in changes that happen in the environment that surrounds them or in the people that participate. This type of experience is difficult to emulate in the real world because it is not easy to move individuals to totally different contexts. This is even more challenging for online classes where it will be impossible to bring everyone to a single location. The more that students participate in this type of activity the more likely that their senses will be sharpened and they will be able to more easily adapt to many different types of circumstances. In the context of the policy simulation the students are putting themselves in a virtual world where they are experiencing real time interaction with their peers who are each playing roles. Their behavior in a synchronous environment will thus differ from the behavior that they are accustomed to from bulletin boards. In this setting they have to spend some time learning about the interests of the people that they need to lobby. They need to be able to identify them as each of the participants is represented in the virtual world as an

avatar. Thus they need to observe all of the characters carefully to be able to identify the person that they want to influence. Much of the learning is not dictated but rather comes from participating in the game.

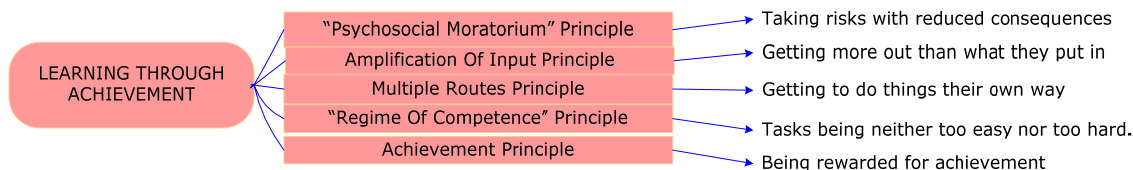


Transfer of Knowledge. While the students learn from the context and the characters that participate in the game the most important element of the game is the possibility of being able to transfer the knowledge that they learn from the virtual environment to real life situation. In the case of the policy simulation one of the objectives of the game is not necessarily to learn the skills of a lobbyist. It is true that part of the content that they need to learn for the information policy class is the different theories about influence but this type of information is not relevant to them unless they can find a way of applying it. Some fields have direct connections with policy while others are indirect. It will be easier for students in fields with stronger policy connections to apply what they learn.



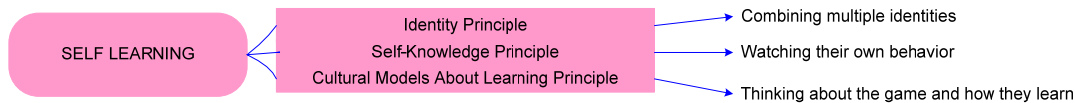
Learning through Practice. It is clear that many kids are highly skilled at playing video games. This happens because they have the opportunity to practice countless times which eventually improves their skills at whatever task they need to complete. Virtual worlds have the potential of providing this type of practice but it is much more limited within the context of a class because it will require the presence of other students. The possibility for practice will depend on how open the teacher wants to make the environment. In order to provide students with opportunities for practice there will need to be a critical mass of participants, all of whom know the instructions. While this could be possible within the context of a commercial virtual world it is less likely to be the case for an educational game. One of the things that educators should keep in mind when developing or modifying games for these virtual worlds is the graduated experience that the participant gets when playing the games. These games begin with simple tasks and then over time the tasks become more complex. It is for this reason that young people return again and again to play because they know that they can, with practice, reach a higher level. With a higher level there is also the reward of finding new things, new experiences, and new challenges. This should be taken into account when setting up a game for a virtual world. Easier tasks should come first and then adding modules or changing the rules to add more elements of greater complexity can potentially keep the students interested in the game to be enthusiastic about playing.

The lobbying simulation that was presented in this paper was not designed in this manner. It was set to be played once with a specific set of rules. The game could be modified to have, for example, a session where the lobbyists organize and then a session with potential policy makers. This can at least provide more than one opportunity to interact with the content of the class as well as providing two different experiences. Of the learning principles identified by Gee the ones related to practice are perhaps the most difficult to achieve within the context of a game in a virtual world.



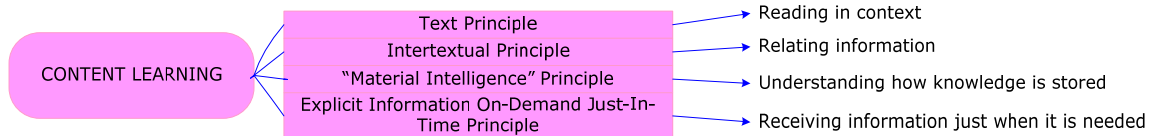
Learning through Achievement. People are limited in their tolerance of disappointment. It has been shown that most people react more positively to positive than to negative feedback. (McKeachie & Gibbs, 2002) Games have the power of providing players with the satisfaction of winning. While in a multi-player game someone will have to lose this is generally understood not as a failure but as an attempt. Young people are thus not discouraged and instead try a new strategy, taking risks. Given that failure in games does not translate into failure in real life the benefit of taking risks outweighs the costs of losing. It is

precisely because of this that players have the opportunity of learning from the many attempts that they make in these worlds. The policy simulation in Habbo Hotel allows the students to try multiple strategies to try to influence their peers. They can use scare tactics, provide statistical information that supports their point, make deals to trade votes, or work with others who have similar interests to try to influence the policy issues at stake. If, for example, they failed to influence one person they may try a different technique the next time and, with trial and error, be able to determine the strategies that work best. Success in a simple, controlled, and fun environment keeps students interested in a task. Having achieved success in easier tasks motivates them to try more complicated ones. Educators, when developing games or simulations, should think about the opportunities that they offer students to achieve success to motivate them to take on more difficult content or skills.



Self Learning. Virtual worlds provide opportunities for people to learn about themselves and about others. Educators have to keep in mind that virtual worlds are not bound by the laws of nature and they are not as bound by traditional social rules. With this in mind games for these worlds can be developed with great creativity and in a fantastic world. Depending on the subject matter educators can ask students to take on real or fictional characters for whom they will then take on different behavioral traits. In this process of assuming somebody else's persona they can learn much about the circumstances that the person faces and the problems or benefits of having such a life. This can then give them an opportunity to compare the situation of this fictional character with their own life and learn from having or facing different circumstances. In the lobbying game the students are given the opportunity to be politicians, librarians, CEOs, presidents and in these roles they will then have to learn about each of these characters' interests as well as the interests of others to be able to be effective at lobbying for their cause.

Evidence of the type of understanding that can take place between people when they take different characters is exemplified by Eric Brown and Asi Burak's game *Peacemaker* which won the Public Diplomacy and Virtual Worlds competition organized by the University of Southern California's Center on Public Diplomacy. In the game the participants take either Israeli or Palestinian characters and the objective is to make peace. This sort of simulation could involve roles such as a CEO, a youth counselor, or a terrorist. By taking somebody else's identity, students can learn about content and context that in the past were only available in books. They can form opinions about a situation not from the detached and foreign context of books but from actually living the experience itself within a virtual world. These experiences also help them form their own views and learn about their own values and interests.

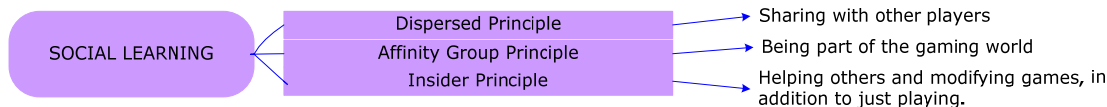


Content. The quality of content is critical for educators. While lectures and discussion boards in an online environment allow students to talk about an issue and reflect on the issue at hand, this knowledge can be superficial and likely to be forgotten. In a virtual world students are put in the middle of situations in a unique context. As a result, they are more likely to remember the issues illustrated given that they will inevitably have to make more effort in preparing and in testing many alternative actions. The mistakes they make and the successes they have will be remembered longer. Their reading now has a more concrete, even if fictional context. They can now decide the type of information that is more relevant for the situation. They are given the opportunity to select information on an as needed basis and take and relate information that is important for the type of activity that they are engaged in. This process thus allows for a much deeper level of knowledge.

The scope of topics that can be explored is nearly unlimited in Second Life. For example, people can develop businesses, pay for items with virtual world currency, and develop real state. Universities and other organizations can have a virtual presence. Virtual worlds are not limited by the physics and rules of the real world and thus can create more memorable situations. In addition, real world situations can be emulated in a virtual world. This flexibility of the medium can allow educators to convey a wide variety of information.

These activities are, however, time consuming to create, a major disincentive for educators. It is much easier to lecture for two hours on a subject than to develop a game, write clear instructions, see that students go to a virtual world and develop their characters, and then hope that the technology will work as expected. It is much more labor intensive for the same amount of knowledge conveyed. In the lobbying simulation, much preparation had to be done ahead of time and the students were required to participate in an orientation session before they were actually able to participate in the game. In order to minimize the amount of time necessary to play an educational game the instructor should try to think of several concepts that can be bundled together such that one effort is able to convey more knowledge. Alternatively one game can be built in a modular manner such as content

experienced at various times in an incremental way. Bundling and gradualism are thus two important components for an effective delivery of content in a virtual world activity.



Social Learning. Prior to data networks video games were solitary and players interacted with the computer alone. With the advent of the Internet and the expansion of broadband, games have increasingly involved interaction among human players. Virtual worlds in their massively multiplayer form could not have existed without the Internet because the nature of the experience relies on the existence of other participants. Educators can take advantage of the social aspect of the game to foster collaboration. A game, although generally conceived as a zero sum experience, can be developed so that effective collaboration rather than direct competition leads to winning. In the lobbying game, social interaction is central to the success of the simulation. Communication allows the participants to practice their skills at influence. In this case winners of the game were those that were able to convert more individuals to get a bill to be voted in their favor. This type of game can create intellectual conflicts because a participant can be required to argue for and vote for a policy that he/she disapproves of.

Challenges for Educators Using Games in Virtual Worlds

While there are many potential rewards, as described in the previous section, there are also important challenges that educators need to keep in mind when developing games for virtual worlds. We were able to identify these challenges from the process of developing the lobbying game.

The first and greatest challenge is to develop the game itself. Most university faculty do not have formal training in education, much less game development. However educators should realize that in the Internet era it should not be difficult for them to find games designed for classroom use that could be adapted for a virtual world. This can substantially reduce the development time.

Second, while there are games already developed, these have not been archived in a single location and thus it will take some time and effort to find them.

Third, the adaptation process is still time consuming because detailed instructions need to be written for the students as a group as well as specific instructions for the participating characters. To make sure that the game works the professor should test the game a couple of times with a small group of students to find weaknesses before using it in a class.

Fourth, technology is unpredictable. Even in the process of developing the game without yet having it tested with students we experienced technological difficulties. In our case the authors could not enter the virtual world by mistake because one of us was entering the American site while the other had prepared the lobbying room from the English site. While this is a minor mistake other technical problems can easily happen, in particular regarding speed of connectivity and learning to navigate and function in a virtual world. To minimize the technical problems it will be desirable, as suggested before, that the professor familiarize the students with the virtual world before running the simulation.

Fifth, anonymity poses difficulties. In entertainment-based virtual worlds anonymity can lead to negative behaviors that, if present in an educational setting, can severely affect the dynamics of the game. In an anonymous environment the participant may take more risks but these can be undesirable risks from the perspective of others. Educators thus have to consider the behavioral standards that would be acceptable in the virtual world while the game is taking place. The teacher should consider when to reprimand and even expel a student who engages in inappropriate behavior. In all of these situations testing the game before using it in a class will help to minimize these challenges.

Future Trends

Virtual worlds are still new environments that are likely to evolve into complex settings that can provide rich and challenging experiences for students. Online distance education is based on the traditional campus education. However because it is no longer necessary to be physically present in a campus there is thus no need to have the student admitted to a particular university to be able to take a class. It is thus possible that in the future we could see models where students can develop their classes a la carte and be able to select from a variety of topics from different institutions and obtain a customized degree.

Virtual worlds could become the online classroom of the future in place of traditional bulletin boards. In these settings the professor can immerse the student in the context that relates to the content of the class. These experiential learning encounters can also be recorded and serve as reference and future learning for students that were unable to attend.

In virtual worlds distance does not matter and thus students can moved from one situation to another in a matter of seconds. Different topics covered in the curriculum could thus take place in multiple contexts. Students which were situated in a classroom alone or on a bulletin board alone could now be taken to multiple settings for many of the topics of the class.

With the popularity of educational games in virtual worlds open source efforts by faculty and students will help to develop more simulations. These worlds will allow developers to work together on the development of a simulated environment and learn in the process as well. In addition, because of these experiences, the learning curve that we experience today will be reduced because these setting will become second nature to students and extensive tutorials prior to games will not be necessary.

In virtual worlds students do not need to take only human personae but also any other organisms. Thus a student learning about zoology can become an insect for example and experience first hand the life of this organism. A chemistry student can become a molecule and once again see first hand how they interact with other elements.

In the context of life long learning students who are about to enter college can get to know what the life of a stock broker, or an archeologist, or a scientist is like. These simulations can move high school students from shadow career days that are difficult to coordinate and expensive to implement to virtual worlds where they can actually experience the careers that they are considering.

At the edges, one of the most exciting parts of virtual worlds can occur when the simulation is no longer a game and becomes a reality. Real world interactions and economic transactions take place. A class that started with a simulation and a role playing game can enter the real world through a virtual experience. Imagine, for example, a psychology student who is learning to become a therapist in a role playing game. He could initially have patients through role play but after gaining some experience they could provide counseling online to people who are unable to attend a therapy session in person. A student can thus learn in a controlled environment prior to encountering patients.

Conclusion

There are several lessons that can be learned from games in virtual worlds. First it is clear that the medium provides students many opportunities to learn about themselves through taking on multiple personae or even other entities. They can learn from participating in many contexts, each of which can provide different content. A multimedia environment in a digitally literate society can help to enrich the experience making the learning much more immersive and compelling. The social interactions can also support learning by allowing collaborations.

These virtual worlds will become the online education of the future because of the flexibility and the richness that educational games can provide students. At this point in time however there are important challenges to overcome. These can be attributed to the fact that we are just beginning to utilize this medium for distance education and we do not yet have the tools and the skills to develop games and work with students. Even though students today are more technologically sophisticated most have not yet been exposed to this type of experience. Thus the learning curve for them is still steep.

Few faculty members are sufficiently aware of these simulated environments and thus few resources have been provided to facilitate this type of education. However, the increasing popularity of these worlds and the presence of some universities and companies inside them will create the critical mass necessary to make development of the environment much easier, such as the creation of libraries that collect virtual world games and simulations for the many disciplines that are taught today in our campuses. The physical worlds will no longer dictate the rules of the virtual worlds and over time a completely different educational experience can become the distance education of the future.

References

- Bartle, R. (1990). Interactive Multi-User Computer Games. MUSE Ltd, British Telecom plc. Retrieved June 30, 2007, from the World Wide Web: <http://www.skepticfiles.org/cowtext/comput~1/mudrepor.htm>
- De Vita, G. (2001). Learning Styles, Culture and Inclusive Instruction in the Multicultural Classroom: A Business and Management Perspective. *Innovation in Education and Training International*, 38, 165-174.
- Fleming, M. J., & Rickwood, D. J. (2001). Effects of violent versus nonviolent video games on children's arousal, aggressive mood, and positive mood. *Journal of Applied Social Psychology*, 31(10), 2047-2071.
- Franklin, S., Peat, M., & Lewis, A. (2003). Non-traditional interventions to stimulate discussion: the use of games and puzzles. *Journal of Biological Education*, 37(2), 79-84.
- Gee, J. P. (2003). *What Video Games Have to Teach us About Learning and Literacy* (First ed.). New York, NY: Pelgrave MacMillan.
- Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology*, 17, 300-312.
- Heffler, B. (2001). Individual Learning Style and the Learning Style Inventory. *Educational Studies*, 27, 307-316.
- James, K. (2006). BT: Real Benefits in Virtual Worlds. A New Start. Retrieved June 30th, 2007, from the World Wide Web: http://xinkaishi.typepad.com/a_new_start/2006/12/bt_real_benefit.html
- Jayakanthan, R. (2002). Application of computer games in the field of education. *Electronic Library*, 20(2), 98-102.
- Kerawalla, L., & Crook, C. (2002). Children's computer use at home and at school: context and continuity. *British Educational Research Journal*, 28(6), 751-771.

- Kusunoki, F., Sugimoto, M., & Hashizume, H. (2000). Towards the integration of physical and virtual worlds for supporting group learning, *Digital Cities* (Vol. 1765, pp. 224-235).
- Laver, M. (1979). *Playing Politics*. New York: Penguin.
- Maher, C., & Corbit, M. (2002). Creating genetic applications for informal science learning in multi-user virtual environments. *Journal of Network and Computer Applications*, 25(4), 295-308.
- Mantovani, F., Castelnovo, G., Gaggioli, A., & Riva, G. (2003). Virtual reality training for health-care professionals. *Cyberpsychology & Behavior*, 6(4), 389-395.
- McKeachie, W., & Gibbs, G. (2002). *Teaching Tips: Strategies, Research, and Theory for College and University Teachers*. Boston, MA: Houghton Mifflin Company.
- Mumtaz, S. (2001). Children's enjoyment and perception of computer use in the home and the school. *Computers & Education*, 36(4), 347-362.
- Subramanian, R., Khang, G. N., & Sai, C. L. (1999). Word Juxtapoz-an Innovative Tool for Promoting Interest in Biological Education. *Journal of Biological Education*, 33, 103-104.

Key Terms and Their Definitions

Distance education: A learning environment for students that take classes primarily on line from different locations “Teaching and learning in which learning normally occurs in a different place from teaching.”

(<http://www1.worldbank.org/disted/glossary.html>).

Course management systems: Software applications designed to facilitate online education using a web interface. The system can manage enrollment, content, and grades among other things.

Lifelong education: The process of continued learning after an individual leaves formal education. This can include college continuing education courses, correspondence, and certifications in physical or distance format.

Video Game: Entertainment software with predetermined goals that can be played through specialized equipment or with a computer.

Virtual World: A graphically immersive, persistent, shared, and typically avatar-based digital environment that is generally hosted on servers connected to the Internet.

Avatar: A graphical image that represents a user in a virtual world.

Simulation: A simplified model that attempts to represent a situation that the creator intends.

Role playing game: A game where the participants take a real or fictional persona for entertainment or education.

FULL PAPER SUBMISSION 13:

Prim Drift, CopyBots, and Folk Preservation: Three Copyright Parables about Art in the Digital Age

AUTHOR(S):

Kraus, Kari M.

ORGANIZATION(S):

University of Maryland, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Prim Drift, CopyBots, and Folk Preservation: Three Copyright Parables about Art in the Digital Age

[FULL PAPER SUBMISSION type: paper proposal]

In 2006 the American Council of Learned Societies published “Our Cultural Commonwealth,” a landmark report on cyberinfrastructure for the humanities and social sciences. As used by the Commission, the term “cyberinfrastructure” covers the aggregate social, technological, institutional, legal, economic, and human capital needed to advance the humanities and social sciences in a global digital environment (ACLS 6). Brett Bobley of the National Endowment for the Humanities provided some indication of the stature of the ACLS report at a recent talk when he referred to it as the “Bible” of the NEH’s new digital humanities initiative.

“Our Cultural Commonwealth” identifies six key challenges that must be overcome if we are to fulfill the promise of a robust cyberinfrastructure, of which arguably the most daunting and complex is copyright (ACLS 18-25). In its 51-page document, the Commission invokes intellectual property in one form or another no less than 47 times, ultimately concluding that although it “can offer no simple solutions” (ACLS 31), the various stakeholders named in the report nonetheless have an obligation to do what they can to strengthen and support the public domain and fair use provisions of the law. To decline this advocacy role is to jeopardize the precious few limitations on absolute copyright upon which we as information professionals rely in our efforts to preserve and transmit the cultural record. Perhaps nowhere is that danger more ominously expressed than in Susan Bielstein’s *Permissions, A Survival Guide: Blunt Talk about Art as Intellectual Property*:

Asking permission [of a rights holder] zealously and unnecessarily also catches you up in a mentality of acquiescence. Acquiescence is a wasting disease rooted in anxiety and ignorance, and it helps propel the all-consuming permissions culture . . . in the quotidian world of intellectual property, acquiescence operates far beneath the beacon eye of statute or treaty, and capitalizing on it is not good for anyone’s health. It wastes time, it wastes money, and it produces a compliant society vulnerable to abuse and wholesale ideological shifts in the law. (10-11)

The purpose of this talk, then, is twofold: first, to furnish a case file of digital art on the web that dramatizes the conflict between private and public rights in creative culture (i.e., the provocative “prim drift,” “copybots,” and “folk preservation” mentioned in my title, which are part of a larger conversation about emergent art genres and creative communication technologies, such as 3D virtual worlds, that, although still in their infancy, will shortly become as established in their own way as television and video are today); and second, to propose, as an experimental intervention into the legal system (and modeled in part after the Electronic Frontier Foundation) a humanities center staffed by archivists, librarians, visual resource curators, art historians, humanities researchers, and lawyers dedicated to testing in a court of law (with the aim of legally validating) claims that, in the words of Gary Schwartz, have the potential to “maximize our own collective position as holders of copyright” in public domain works or, I would add, as users of copyright in private domain works.

The choice of an arts and humanities dataset to seed the discussion about copyright is inspired in part by a venerable roster of artists who have challenged prevailing assumptions about the nature of creativity, originality, and authorship. These assumptions in turn inform cultural attitudes toward intellectual property. The cut-ups of William S. Burroughs; the ready-mades of Marcel Duchamp; the appropriation art of Sherrie Levine, Jeff Koons, and Andy Warhol all potentially run afoul of modern IP law. For this reason, the arts frequently act as an early warning system for copyright, alerting us to potential ambiguities in the case law or to legislative actions that alter, intentionally or not, the delicate balance between the rights of creators and the rights of users that the framers of the Constitution sought to protect. It is not only artists themselves, but also the collectors, scholars, publishers, museums, libraries, executors, and rights agents drawn into their orbit who have the power to upset or restore the golden mean of copyright.

The “parables” mentioned in my title are central to this discussion. Parables, of course, are stories with messages. This talk will include three such stories, of which the first, *Prim Drift*, is previewed below. The other two (*CopyBots* and *Folk Preservation*) will be developed in the full version of the presentation.

In the multi-user virtual environment *Second Life*, “prims”—short for “primitives”—are the basic building blocks from which 3D objects such as cars, houses, books, and furniture are built. Individual prims can be linked together into larger systems to form complex shapes. Although generally stable, such builds will sometimes “drift” over time: like tectonic plates sliding past each other, prims can randomly shift out of position, producing distorted visual effects (Walsh, “Drifting Apart”). While content creators can edit their scripts to repair the damage, potential problems arise once the items have been sold. This is because every creative asset in SL is encoded with a set of permissions that dictate how it may be distributed and used by other residents. The permissions system includes four fields: move, modify, copy, and transfer. If either the second or fourth of these is disabled, the owner of the object, as distinct from its creator, is prevented from exercising freedoms that are normally either unregulated or protected by law in real life: the right to modify an object for personal use and the right to resell or give away a lawfully obtained copy under the First Sale Doctrine. Thus, for example, in RL I may annotate my copy of *Gone with the Wind* in preparation for class discussion or take it to the bookbinder for repair if the pages come loose or donate it to my local public library when I no longer need or want it without infringing on the author’s copyrights. By contrast, if I own a book in SL, I may not be able to do these things: I may not be able to repair the book if it shows signs of prim drift or give it as a gift to someone else. The inability to alter objects for personal use is particularly troublesome: it is as if I suddenly found myself in a world where there were no tailors to hem pants or cobblers to mend shoes or antique experts to restore Grandma’s heirloom hutch.

The set of problems I'm describing are as applicable to the 2D web as they are to the 3D web. Increasingly contract law overrides copyright law in electronic commerce. End User Licensing Agreements (EULAs), for example, complicate the transfer of software applications from one party to another, suggesting that the erosion of the First Sale and Fair Use doctrines in Second Life is merely part of a larger trend.

Because this example and the other two mentioned in my title function as limit cases for copyright policy, showing us where the system breaks down in a digital environment and allowing us to probe its edges, they help justify what might otherwise be seen as an implausible proposal: to establish a humanities center that would expand the site of activism to include not only the customary conference proceedings, working groups, white papers, articles, and symposia, but also, crucially, the courts. Nearly a decade ago, art historian Gary Schwartz urged his colleagues in a spirited talk at the annual meeting of the College Art Association to adopt more assertive measures to revive an ailing fair use doctrine. In an attempt to persuade his audience members of the need for innovative approaches, Schwartz informed them that he had been "assured by counsel that test trials on any of these [copyright] issues, in a well-chosen court, stand an excellent chance of success." In 1998, such a strategy might have seemed improbable; in 2008, I'm convinced it is not. With the establishment of a grant initiative in 2003 to fund work "in the area of intellectual property and the long-term protection of the public domain," the MacArthur Foundation has signaled its readiness to support experimental programs; recent recipients include the Electronic Frontier Foundation, which was awarded over half a million dollars in 2007. The EFF--whose mandate is to protect the public interest in a networked world--helps defend free speech, privacy, and consumer rights, in part by subjecting them to legal proceedings in a court of law. It thus serves as an important precedent for the course of action advocated here.

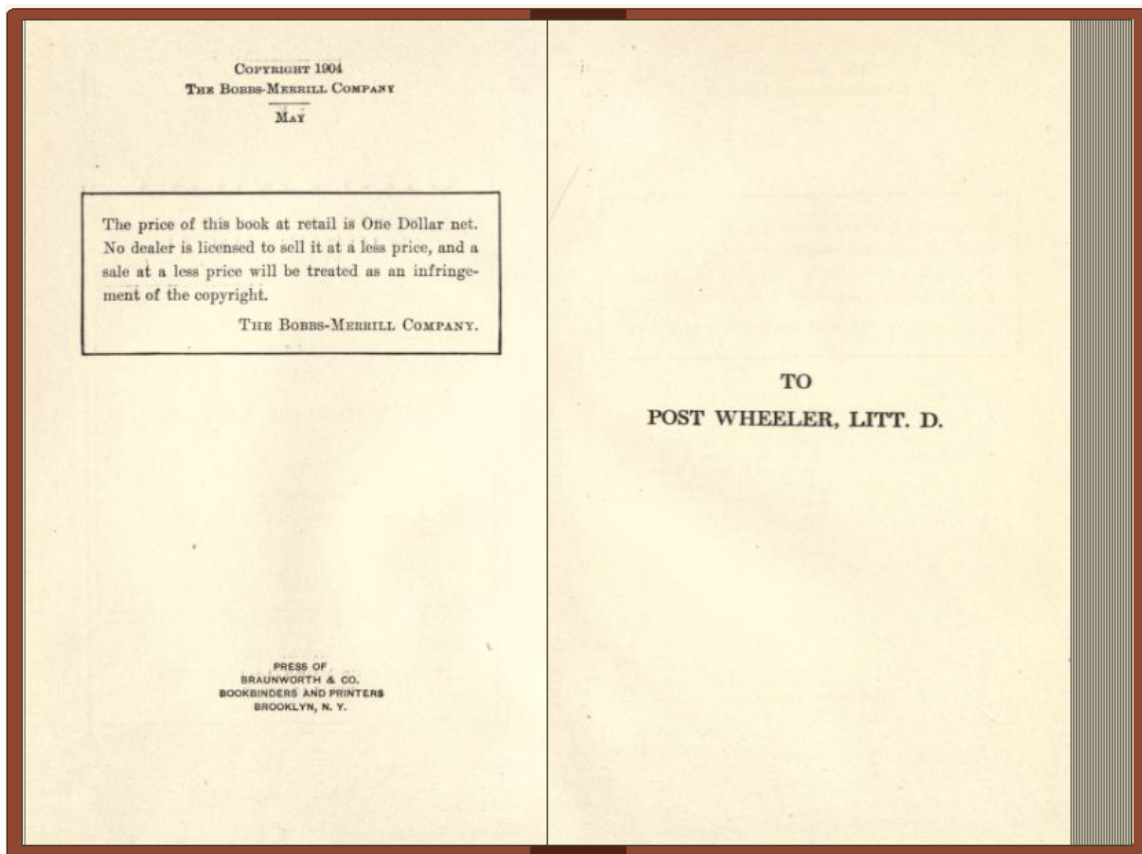


Figure 1: Notice (left page) published by the Bobbs-Merrill Co. in *The Castaway* by Hallie Erminie Rives in 1904. The attempt to use the right of distribution to control resale of the book eventually led to the landmark Supreme Court decision *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339 (1908). The case codified The First Sale Doctrine.



Figure 2: A physical copy of the novel *Anima* by Dalian Hansen (digitalKu 2007). The freedoms to modify (write in the margins, dogear the pages, etc.) and resell lawfully obtained copies of the book are protected by Fair Use, First Sale, and de minimis doctrines.

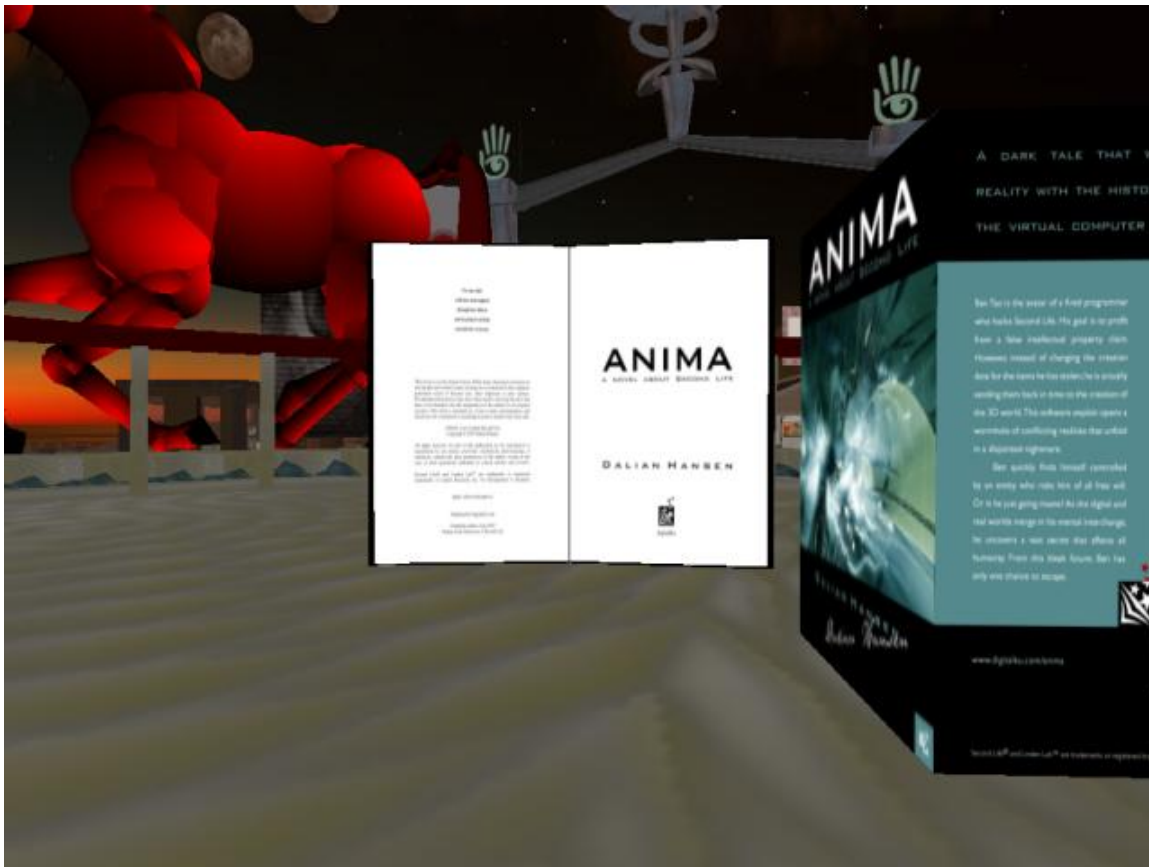


Figure 3: Virtual Copy of *Anima* by Dalian Hansen. This version of the novel can only be read in the virtual world of Second Life. In this case the author has enabled transfer rights, but not modification rights. The permissions system of SL partially nullifies the Fair Use, First Sale, and de minimis doctrines that operate in real life.

ACLS, "Our Cultural Commonwealth." New York: ACLS, 2006.

Bielstein, Susan M. Permissions, A Survival Guide: Blunt Talk about Art as Intellectual Property. University Of Chicago Press, 2006.

Bobley, Brett. "A Candid Chat About the NEH's Digital Humanities Initiative." MITH Digital Dialogue. University of Maryland. 16 October 2007.

Electronic Frontier Foundation. 29 Oct 2007 <http://www.eff.org>.

MacArthur Foundation. "Intellectual Property and the Public Domain." 29 Oct 2007
 <http://www.macfound.org/site/c.lkLXJ8MQKrH/b.943331/k.DA6/General_Grantmaking__Intellectual_Property.htm>.

---. "Recent Grants." 29 Oct 2007
http://www.macfound.org/site/c.lkLXJ8MQKrH/b.1002695/k.9993/General_Grantmaking__Intellectual_Property__Recent_Grants.htm.

Schwartz, Gary. "No Fair: Long-Term Prospects for Regaining Unencumbered Use." Toronto Town Meeting Program. College Art Association. February 1998. 29 Oct 2007 <<http://www.pipeline.com/~rabaron/ttm/SCHWARTZ.htm>>.

Walsh, Tony. "Drifting Apart." Clickable Culture. 10 June 2004. 29 Oct 2007
 <http://www.secretlair.com/index.php?/clickableculture/entry/my_second_life_part_9/>.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)
 Prim Drift, CopyBots, and Folk Preservation: Three Copyright Parables about Art in the Digital Age

Kari Kraus
 University of Maryland
 College of Information Studies
 Department of English
 College Park, MD 20742-4345
 1.240.997.9759
kkraus@umd.edu

ABSTRACT

This paper employs a series of case studies from the domains of digital arts and creative/experimental new media to elicit tensions and contradictions in the current state of copyright and intellectual property law. I pay particular attention to the role of the "pirate" as preservationist--rather than taint or corrupt, historically we know that piracy has helped guarantee the survival of important works of literature and art. Throughout, I insist that the humanist is not a dabbler or interloper in these matters; humanistic knowledge, particularly semiotics (the study of sign systems) has the potential to lend consistency and coherence to case law that is currently shot through with loopholes, contradictions, and dead ends. To that end, I also outline the potential of a center devoted to intellectual property law and humanities advocacy.

Topics

Information policy, ethics, and law
 Preserving digital information

Keywords

Second Life, virtual worlds, intellectual property, copyright, arts and humanities, preservation, multi-user virtual environments

15. INTRODUCTION

In 2006 the American Council of Learned Societies published "Our Cultural Commonwealth," a landmark report on cyberinfrastructure for the humanities and social sciences. As used by the Commission, the term "cyberinfrastructure" covers the aggregate social, technological, institutional, legal, economic, and human capital needed to advance the humanities and social sciences in a global digital environment

(ACLS 6). Brett Bobley of the National Endowment for the Humanities provided some indication of the stature of the ACLS report at a recent talk at the Maryland Institute for Technology in the Humanities when he referred to it as the “Bible” of the NEH’s new digital humanities initiative.

“Our Cultural Commonwealth” identifies six key challenges that must be overcome if we are to fulfill the promise of a robust cyberinfrastructure, of which arguably the most daunting and complex is copyright (18-25). In its 51-page document, the Commission invokes intellectual property no less than 47 times, ultimately concluding that although it “can offer no simple solutions” (31), the various stakeholders named in the report nonetheless have an obligation to do what they can to strengthen and support the public domain and fair use provisions of the law. To decline this advocacy role is to jeopardize the precious few limitations on absolute copyright upon which we as information professionals rely in our efforts to preserve and transmit the cultural record. Perhaps nowhere is that danger more ominously expressed than in Susan Bielstein’s *Permissions, A Survival Guide: Blunt Talk about Art as Intellectual Property*:

Asking permission [of a rights holder] zealously and unnecessarily also catches you up in a mentality of acquiescence. Acquiescence is a wasting disease rooted in anxiety and ignorance, and it helps propel the all-consuming permissions culture . . . in the quotidian world of intellectual property, acquiescence operates far beneath the beacon eye of statute or treaty, and capitalizing on it is not good for anyone’s health. It wastes time, it wastes money, and it produces a compliant society vulnerable to abuse and wholesale ideological shifts in the law. (10-11)

The purpose of this talk, then, is twofold: first, to furnish a case file of network-driven digital art that dramatizes the conflict between private and public rights in creative culture (i.e., the provocative “prim drift,” “copybots,” and “folk preservation” mentioned in my title, which are part of a larger conversation about emergent art genres and creative communication technologies, such as 3D virtual worlds, that, although still in their infancy, will shortly become as established in their own way as television and video are

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse

today); and second, to propose, as an experimental intervention into the legal system (and modeled in part after the Electronic Frontier Foundation) a humanities center staffed by archivists, librarians, visual resource curators, art historians, humanities researchers, and lawyers dedicated to testing in a court of law (with the aim of legally validating) claims that, in the words of Gary Schwartz, have the potential to “maximize our own collective position as holders of copyright” in public domain works or, I would add, as users of copyright in private domain works.

The choice of an arts and humanities dataset to seed the discussion about copyright is inspired in part by a venerable roster of artists who have challenged prevailing assumptions about the nature of creativity, originality, and authorship. These assumptions in turn inform cultural attitudes toward intellectual property. The cut-ups of William S. Burroughs; the ready-mades of Marcel Duchamp; the appropriation art of Sherrie Levine, Jeff Koons, and Andy Warhol all potentially run afoul of modern IP law. For this reason, the arts frequently act as an early warning system for copyright, alerting us to potential ambiguities in the case law or to legislative actions that alter, intentionally or not, the delicate balance between the rights of creators and the rights of users that the framers of the Constitution sought to protect. It is not only artists themselves, but also the collectors, scholars, publishers, museums, libraries, executors, and rights agents drawn into their orbit who have the power to upset or restore the golden mean of copyright.

16. PRIM DRIFT

In the multi-user virtual environment Second Life, “prims”—short for “primitives”—are the basic building blocks from which 3D objects such as cars, houses, books, and furniture are built. Individual prims can be linked together into larger systems to form complex shapes. Although generally stable, such builds will sometimes “drift” over time: like tectonic plates sliding past each other, prims can randomly shift out of position, producing distorted visual effects (Walsh, “Drifting Apart”). While content creators can edit their scripts to repair the damage, potential problems arise once the items have been sold. This is because every creative asset in SL is encoded with a set of permissions that dictate how it may be distributed and used by other residents. The permissions system includes four fields: move, modify, copy, and transfer. If either the second or fourth of these is disabled, the owner of the object, as distinct from its creator, is prevented from exercising freedoms that are normally either unregulated or protected by law in real life: the right to modify an object for personal use and the right to resell or give away a lawfully obtained copy under the First Sale Doctrine (Figures 1, 2, and 3). Thus, for example, in RL I may annotate my copy of *Gone with the Wind* in preparation for class discussion or take it to the bookbinder for repair if the pages come loose or donate it to my local public library when I no longer need or want it without infringing on the author’s copyrights. By contrast, if I own a book in SL, I may not be able to do these

things: I may not be able to repair the book if it shows signs of prim drift or give it as a gift to someone else. The inability to alter objects for personal use warrants further consideration: as an analogy, it would be as if I suddenly found myself in a world where there were no tailors to hem pants or cobblers to mend shoes or antique experts to restore Grandma's heirloom hutch.

The set of problems I'm describing are as applicable to the 2D web as they are to the 3D web. Increasingly contract law overrides copyright law in electronic commerce. End User Licensing Agreements (EULAs), for example, complicate the transfer of software applications from one party to another, suggesting that the erosion of the First Sale and Fair Use doctrines in Second Life is part of a much larger trend.

Because this example and the other two mentioned in my title function as limit cases for copyright policy, showing us where the system breaks down in a digital environment and allowing us to probe its edges, they help justify what might otherwise be seen as an implausible proposal: to establish a humanities center that would expand the site of activism to include not only the customary conference proceedings, working groups, white papers, articles, and symposia, but also, crucially, the courts. Nearly a decade ago, art historian Gary Schwartz urged his colleagues in a spirited talk at the annual meeting of the College Art Association to adopt more assertive measures to revive an ailing fair use doctrine. In an attempt to persuade his audience members of the need for innovative approaches, Schwartz informed them that he had been "assured by counsel that test trials on any of these [copyright] issues, in a well-chosen court, stand an excellent chance of success." In 1998, such a strategy might have seemed improbable; in 2008, I'm convinced it is not. With the establishment of a grant initiative in 2003 to fund work "in the area of intellectual property and the long-term protection of the public domain," the MacArthur Foundation has signaled its readiness to support experimental programs; recent recipients include the Electronic Frontier Foundation, which was awarded over half a million dollars in 2007. The EFF--whose mandate is to protect the public interest in a networked world--helps defend free speech, privacy, and consumer rights, in part by subjecting them to legal proceedings in a court of law. It thus serves as an important precedent for the course of action advocated here.

What would distinguish the center I envision from either the EFF or Stanford's Center for Internet and Society is the role of the humanist, who would serve not as an ancillary figure, but one central to the center's mission and success. Although I suspect many are capable of envisioning humanities scholars as evangelists, public relations spokespeople, legal assistants, legal researchers, or legislative advocates in this enterprise, I believe they are qualified to do more: to serve as expert witnesses or consultants in civil cases involving intellectual property and the arts. The literary scholar who specializes in textual transmission, the philosopher who investigates

theories of language or the nature of universals, and the art historian who publishes on reproductive printmaking are all de facto semioticians, experts in sign systems and how they work. Because intellectual property concerns itself with originals and copies, a substantial body of case law hinges on the semiotic distinction between the two. The ability to state when two objects are fundamentally the same or different--to determine, for example, if a copy of Salvador Dali's *Persistence of Memory* is primarily substitutive or transformative--is therefore a legal desideratum, one which many humanists are qualified to help address.

"Given the central and fundamental role of sign systems," writes Jeffrey Long in a special issue of *Semiotica*, "why is the field that studies them--semiotics--so unknown amongst many who wish to solve practical problems in business, science, the arts . . . government" and, we might add, the law (1)? A preliminary survey of judicial opinions on court cases that involve accusations of copyright infringement of visual works of art turns up a network of seemingly contradictory, ad hoc, or ambiguous rulings that would benefit from the rigorous application of semiotic principles of individuation, principles that would allow us to make tighter discriminations among things like originals, copies, versions, remixes, editions, and other relevant categories. The point here is that semiotics could be instrumental in helping courts adjudicate between the competing copyright claims of plaintiffs and defendants, as well as in fortifying fair use and the public domain. Contrary to the position recently espoused by Stanley Fish, professor of English and Law at Florida International University, regarding the lack of any utilitarian value whatsoever of a 21st century humanities education, the center I'm proposing would provide an affirmation of the power of humanistic knowledge to shape public policy in the digital age.

17. FOLK PRESERVATION

By "folk preservation," I mean preservation that is amateur rather than professional; distributed rather than centralized; and unauthorized rather than authorized. It is often of an avocational rather than vocational nature. Many of the current preservation initiatives in Second Life, for example, are folksonomic in character: one of the longest standing structures in the metaverse, a statue representing the burning man effigy of the annual art festival held in the deserts of Nevada ("The Man"), continues to survive in an old parcel of Natoma, not, at least initially, because of any top-down decree from Linden Lab, the developer of SL, or the intervention of Real Life (RL) or Second Life (SL) conservators, but because of the on-going support and efforts of residents.

Of the many varieties of folk preservation, the one on which I'd like to focus is piracy. Consider three examples: According to industry insiders, the New York Philharmonic was ironically forced some years ago, when it began to think seriously about its preservation program, to purchase

surreptitiously made sound recordings of its live performances from concert-goers who had smuggled their portable recording devices into the concert hall (Manildi; Winner). The 10-CD commemorative archival box set of Leonard Bernstein Live, for example, which was issued by the NY Philharmonic, includes an entire Wagner concert from 1970 taken from an illicit tape made by Roger Frank (Godell). Such bootleg music was a mainstay at William Lerner's record shop Music Masters in the 60s and 70s, a hangout for musicians and collectors often in search of rare material (Kozinn). Although Lerner's reasons for hiring freelancers armed with recorders to fill the seats at concert halls were strictly commercial, both he and Frank nonetheless join the historical ranks of pirates who deserve recognition for archival and preservation achievement in the arts.

My second example comes from the world of cinema. In an online interview, the British screenwriter Peter Briggs relates how he lost and subsequently found his *Alien vs. Predator* script:

I wrote "A vs P" originally . . . on an Amstrad computer, which was about one step above a Univac Room Filler. In '92 I swapped to an Apple Mac, which I've used ever since. And I ended up losing the Amstrad disk, which was some weird, unreadable proprietary brand anyway. It wasn't until whoever it was transcribed it and pirated it onto the web years later, that I was able to cut-and-paste it into Final Draft and have an electronic copy again. So, thank-you, Internet Leaker, wherever you [are]! (Qtd. by Doctorow)

My third example, the one on which I'd like to spend the most time, is *Agrippa* (a book of the dead), an artist's book co-authored by the writer William Gibson and the artist Dennis Ashbaugh, and published by Kevin Begos, Jr. in 1992. Described in contemporary press releases as a "multi-unit artwork" (Gibson, "Introduction to *Agrippa*"), *Agrippa* is difficult to classify, both physically and generically. It was originally sold in two versions. The deluxe version came wrapped in a shroud, its cover artificially aged and its pages scorched—"time-burned," like the photo album described in Gibson's poem ("*Agrippa Files*"). Inside are etchings by Ashbaugh and double columns of DNA that ostensibly encode the genome of a fruitfly (Kirschenbaum xi). Nestled in the center of the book is a 3 1/2 inch floppy disk that contains the poem, a meditation on time, memory, and decay. Its governing metaphor is that of the mechanism "a trope," notes Matthew Kirschenbaum, "that manifests itself as a photograph album, a Kodak camera, a pistol, and a traffic light, as well as in less literal configurations. (ix)" *Agrippa*, among other things, is about our misplaced faith in the permanence and objectivity of media. Like human memories, media distort, invent, and erase the very objects they're

designed to preserve: handwriting fades and becomes illegible, photographs break the fourth wall by constantly reminding us of the world that lies just outside the picture frame, and inert technological artifacts put up no resistance when new ones come along to replace or destroy them. Paradoxically, the speaker of the poem takes recourse in his own recollections to supplement the incomplete records of the past, records that were originally intended to compensate for the limitations of memory. By such a process, he tries to recapture, for instance, the smells of the saw-mill once owned by his father, whose "tumbled boards and offcuts" are pictured in an old photograph (Gibson *Agrippa*). Drawing on a synaesthetic imagination, he uses the visual stimulus to prompt an olfactory memory. This complex interplay between mechanism and memory structures the poem as a whole and shapes its manifold meanings.

Agrippa's core themes are expressed through form as well as content: some of Ashbaugh's etchings are overlaid with images printed in uncured toner, which are inevitably smudged and distort those beneath and facing them when the pages are turned (Kirschenbaum xi). More stunningly, as has been often described, Gibson's electronic poem is encrypted to scroll automatically down the screen once before being irrevocably lost, its text disappearing after a single reading. *Agrippa* is therefore subjected, like all material objects, to the forces of decay, but here those forces are manufactured rather than natural, causing the work to disintegrate at an accelerated rate. Ashbaugh, in particular, took considerable delight in anticipating how this volatility would confound librarians, archivists, and conservators: as Gavin Edwards explains, to register the book's copyright, Ashbaugh "would need to send two copies to the Library of Congress. To classify it, they . . . [would] have to read it, and to read it, they . . . [would have to] destroy it." Significantly, however, it was not the librarians who found a workaround to the problem, but the pirates. On December 9, 1992, a group of New York University students secretly video-recorded a live public performance of *Agrippa* at The Americas Society, an art gallery and experimental performance space in New York City. After transcribing the poem, they uploaded it as a plain ASCII text to MindVox, a notorious NYC Bulletin Board, "the Hells Angels of Cyberspace," according to Wikipedia, where it was readily available for download and quickly proliferated across the web. "The Hack," as the incident has come to be called, is told with the hard-boiled suspense of a detective story by Matthew Kirschenbaum, who uncovered the gritty details while working on his new book *Mechanisms: New Media and the Forensic Imagination*, recently published by MIT P.

Although many rare archival and research materials (images, scans, transcriptions, video footage, and simulations) related to *Agrippa* are published on the *Agrippa Files* website with the permission of Kevin Begos Jr., the complete text of the poem--arguably the central node of the work--is not. The editorial team at the University of California Santa Barbara

provides the following notice by way of explanation:

"Currently, the Agrippa Files does not have permission to reproduce the full text of the poem. Many wild copies of the text exist on the Internet. The official copy is on William Gibson's Web site" ("The Poem"). What is worth noting here is that the access mission of the site is to some degree undercut by copyright laws. Whereas publicly accountable archival projects must abide by these laws, they go unheeded by shadow groups, such as the MindVox users, who transmit culture at risk of civil litigation. The irony, of course, is that despite the poem's omnipresence on the web, it is absent from the one authoritative site dedicated to documenting and preserving it. What the incident underscores is the way in which short-term private rights potentially obstruct long-term public rights in cultural heritage.

18. COPYBOTS

In an entry dated from January 2008 on Architecture +, a blog covering commercial products and services for multi-user virtual environments, JeanRicard Broek cryptically proclaims that a "Library of Congress copybot is preserving virtual worlds." Broek, it turns out, is reporting on a recently funded grant sponsored by the Library of Congress to preserve early computer games, interactive fiction, and 3D virtual worlds. The multi-institutional project, with which I'm directly involved, includes the University of Illinois Urbana-Champaign (Principal Investigator), Stanford University, the University of Maryland, and the Rochester Institute of Technology. Because interactive media are highly complex and at high risk for loss as technologies rapidly become obsolete, the project seeks to provide basic standards for metadata and content representation, and conduct a series of archiving case studies for early video games, electronic literature, and Second Life.

The "CopyBot" to which Broek refers is the famed software tool that wreaked havoc on the fledgling economy of Second Life in 2006. Developed by libsecondlife, which describes itself as a project directed toward "understanding how Second Life works from a technical perspective, and extending and integrating the metaverse with the rest of the web," the program generates clones of in-world objects, such as avatars, buildings, books, furniture, cars, and the like (Reuters). Libsecondlife, which originally created CopyBot with the blessing of Linden Lab, intended for it to be used as a debugging and back-up application (Reuters). Once the source-code was released and distributed through the SL Exchange marketplace, however, it was edited and recompiled to duplicate objects without authorization from content creators, to the great consternation of artists and merchants, many of whom apparently suffered considerable financial losses as a result (Wikipedia, "CopyBot"). Nonetheless, the CopyBot is a technology that, although used by some customers to violate the law, is capable (like the home video recorders, photocopying machines, and computers that came before it) of "substantial non-infringing uses" (EFF, Betamax Case).

What are we to make of Broek's analogy, however facetious, between a legitimate, federally funded preservation program and a software tool best known for its ability to facilitate criminal activity? The easy conflation of preservation with piracy no doubt has something to do with the fact that the simple act of copying is central to both. One of the most reliable ways to preserve an object, after all, is to duplicate it. This idea is so basic that it has been codified as a general principle: "lots of copies keep stuff safe" (LOCKSS). But an intellectual property system that grants only minimal exemptions to libraries, museums, and archives from the prohibition against copying provides insufficient grounds for putting this principle into action—unless, that is, there are official or unofficial lines of communication that exist between them and a grey market economy that plays by different rules.

This, indeed, is the lesson of history: our creative heritage infrastructure has not evolved independently of piratical practices, but co-dependently with them. Writing about early European print culture, for example, in his magisterial *The Nature of the Book*, Adrian Johns observes that in the sixteenth through the eighteenth centuries "pirates were . . . not a distinguishable social group. They existed at all ranks of the Stationers' community, [the guild responsible for enforcing copyrights], and at times were among its most prominent and upstanding members" (167). Similarly, Meredith McGill suggests that unauthorized reprinting in nineteenth-century antebellum America was so pervasive--so fully embedded in everyday publishing practices--that it constituted "part of the horizon of the ordinary" (4). We need, then, to better understand the systems of exchange that exist between lawful and unlawful preservation regimes and determine the extent to which the former is beholden to the latter. And we need to get a better purchase on the intellectual property restrictions that apply, directly or indirectly, to the contact zone between them. (For example, if a benefactor were to donate a large collection of materials to a library that included pirated books or software, the First-Sale Doctrine would presumably be rendered moot, making parts of the transaction illegal).

Giving credence to the historical importance of the pirate as preservationist provokes new kinds of questions, some answerable, some probably not:

- Are there authors whose works have descended to us exclusively in pirated editions or unauthorized copies or reprints? If so, who are they?
- If we could tally the total number of books that exist in the world and divide them into two groups, one containing authorized copies, the other unauthorized, what would be the ratio of one to the other? Will we ever have the means to quantify piracy? The difficulty of developing criteria that would allow us to confidently assign a book to one

category or the other would only serve to underscore some of the issues raised here.

- Is pursuing a kind of heroic preservation, whereby we seek to transmit a cultural artifact wholly intact and unchanged to posterity, a fool's errand? Does cultural piracy demonstrate the truth of the old adage that the perfect is the enemy of the good? Is the transformation of culture from one generation to the next something to be valued rather than lamented? Can a fragment or a shard serve as a synecdoche for a larger whole that once existed?⁴

My background in textual scholarship and bibliography tells me that we would be wise to lower the threshold of what constitutes successful preservation. The distinction between a legitimate preservation regime and a piratical one is often illusory, based on the faulty assumption that the former is the guarantor of stability, while the latter delivers only volatility. As an antidote to such notions, I'd like to close with an account of textual transmission found in the British playwright Tom Stoppard's *The Invention of Love*. Stoppard's play tells the story of the early twentieth-century poet and classicist A. E. Housman, who, having just died, reminisces about his life while being ferried across the river Styx by the boatman Charon in the underworld. In this passage, an Oxford don narrates the reception history of the Latin poet Catullus, whose textual fate is shared in its broad strokes by countless other writers of antiquity. Here, then, is a tale of preservation, blemishes and all:

Anyone with a secretary knows that what Catullus really wrote was already corrupt by the time it was copied twice, which was about the time of the first Roman invasion of Britain: and the earliest copy that has come down to us was written about 1,500 years after that. Think of all those secretaries!—Corruption breeding corruption from papyrus to papyrus, and from the last disintegrating scrolls to the first new-fangled parchment books, with a thousand years of copying-out still to come, running the gauntlet of changing forms of script and spelling, and absence of punctuation—not to mention mildew and rats and fire and flood and Christian disapproval to the brink of extinction as what Catullus really wrote passed from scribe to scribe, this one drunk, that one sleepy, another without scruple, and of those sober, wide-awake, and scrupulous, some ignorant of Latin, and some, even

worse, fancying themselves better Latinists than Catullus—until!—finally and at long last—mangled and tattered like a dog that has fought its way home, there falls across the threshold of the Italian Renaissance the sole surviving witness of thirty generations of carelessness and stupidity: the Verona Codex of Catullus; which was almost immediately lost again, but not before being copied with one last opportunity for error. And there you have the foundation of the poems of Catullus as they went to the printer for the first time, in Venice 400 years ago. (24-25)

19. ACKNOWLEDGEMENTS

I would like to thank Donald Manildi, curator of the International Piano Archives at the University of Maryland, and Seth Winner, sound preservation engineer at the New York Public Library, for providing background on the New York Philharmonic and its preservation program, as well as information on the surreptitious sound recordings made in the 60s and 70s. I would also like to thank Alan Liu, project leader of the Agrippa Files and Professor of English at the University of California, Santa Barbara; and Matthew Kirschenbaum, Associate Director of MITH and Associate Professor of English at the University of Maryland, for sharing their wealth of knowledge about Agrippa. Finally, I'd like to acknowledge the learned members of SHARP-L--a listserv sponsored by the Society for the History of Authorship, Reading, and Print—for their generous responses to my questions about preservation and book piracy. While I didn't have the opportunity to follow-up on some of their valuable leads, readers may wish to consult the public record of the conversation, which begins here: <https://listserv.indiana.edu/cgi-bin/wa-iub.exe?A2=ind0801&L=SHARP-L&T=0&F=&S=&P=8206>.

20. REFERENCES

[Unless otherwise noted, all URLs were last accessed on 25 January 2008.]

The Agrippa Files. Transcriptions Project, University of California Santa Barbara. 21 January 2008
<http://agrippa.english.ucsb.edu/>.

Bielstein, Susan M. *Permissions, A Survival Guide: Blunt Talk about Art as Intellectual Property*. University Of Chicago Press, 2006.

Bobley, Brett. "A Candid Chat About the NEH's Digital Humanities Initiative." MITH Digital Dialogue. University of Maryland. 16 October 2007.

Broek, JeanRicard. "Library of Congress CopyBot is Preserving Second Life." *Architecture* + January 2008

⁴ On the fragment as a unit of preservation, see David Lowenthal, "Material Preservation and its Alternatives," *Perspecta* 25 (1989): 66-77.

<http://jeanricardbroek-architect.blogspot.com/2008/01/library-of-congress-copybot-is.html>.

"CopyBot." Wikipedia, The Free Encyclopedia. 14 Jan 2008, 15:48 UTC. Wikimedia Foundation, Inc. 26 Jan 2008
<http://en.wikipedia.org/w/index.php?title=CopyBot&oldid=184268503>.

Doctorow, Cory. "Alien v Predator Script Saved by Internet Pirates." 6 November 2004 BoingBoing
<http://www.boingboing.net/2004/11/06/alien-v-predator-scr.html>.

Electronic Frontier Foundation. 29 Oct 2007
<http://www EFF.org>.

--. "The Betamax Case."
<http://w2 EFF.org/legal/cases/betamax/>.

Edwards, Gavin. "CyberLit: William Gibson's Latest Story Costs \$450, Comes on Disk, and Self-Destructs after One Reading." Details June 1992. Transcription available at
<http://www.matarese.com/matarese-files/3580/cyberlit-william-gibson-latest-story-costs-comes-disk-self-destructs/index.html>.

Fish, Stanley. "Will the Humanities Save Us?" New York Times 6 January 2008
<http://fish.blogs.nytimes.com/2008/01/06/will-the-humanities-save-us/>.

Gibson, William. Agrippa: A Book of The Dead
<http://www.williamgibsonbooks.com/source/agrippa.asp>.

Godell, Tom. "Bernstein Live." American Record Guide 1 March 2001
http://www.accessmylibrary.com/coms2/summary_0286-10950492_ITM.

Kirschenbaum, Matthew G. Mechanisms: New Media and the Forensic Imagination. Cambridge: MIT Press, 2008.

Kozinn, Alan. "A Music Lovers' Mecca is Closing." New York Times 7 November 1990
<http://query.nytimes.com/gst/fullpage.html?res=9C0CE7DF1E3CF934A35752C1A966958260>.

Johns, Adrian. The Nature of the Book: Print and Knowledge in the Making. Chicago: Chicago UP, 1998.

Libsecondlife.
http://www.libsecondlife.org/wiki/Main_Page.

LOCKSS [Lots of Copies Keep Stuff Safe].
<http://www.lockss.org/lockss/Home>.

Long, Jeffrey. "Editor's Note." Semiotica 125 (1999): 1-13. Special issue on notation systems.

MacArthur Foundation. "Intellectual Property and the Public Domain." 29 Oct 2007
http://www.macfound.org/site/c.lkLXJ8MQKrH/b.943331/k.DA6/General_Grantmaking__Intellectual_Property.htm.

---. "Recent Grants." 29 Oct 2007
http://www.macfound.org/site/c.lkLXJ8MQKrH/b.1002695/k.9993/General_Grantmaking__Intellectual_Property__Recent_Grants.htm.

Reuters, Adam. "Outcry as 'CopyBot' Threatens Copyright Protection." Reuters Second Life News Center 14 November 2006
<http://secondlife.reuters.com/stories/2006/11/14/outcry-as-copybot-threatens-copyright-protection/>.

"The Man." Second Life Wiki
http://secondlife.wikia.com/wiki/The_Man.

McGill, Meredith. American Literature and the Culture of Reprinting, 1834-1853. Philadelphia: U of Philadelphia P, 2003.

Our Cultural Commonwealth: The final report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities & Social Sciences. Washington DC: ACLS, 2006
www.acls.org/cyberinfrastructure/acls.ci.report.pdf.

Schwartz, Gary. "No Fair: Long-Term Prospects for Regaining Unencumbered Use." Toronto Town Meeting Program. College Art Association. February 1998. 29 Oct 2007
<http://www.pipeline.com/~rabaron/ttm/SCHWARTZ.htm>.

Stoppard, Tom. The Invention of Love. New York: Grove P, 1998.

Walsh, Tony. "Another Look at 'Second Life' Copyright Issues." Clickable Culture. 11 September 2006.
http://www.secretlair.com/index.php?/clickableculture/entry/another_look_at_second_life_copyright_issues/.

--. "Drifting Apart." Clickable Culture. 10 June 2004. 29 Oct 2007
http://www.secretlair.com/index.php?/clickableculture/entry/my_second_life_part_9/.

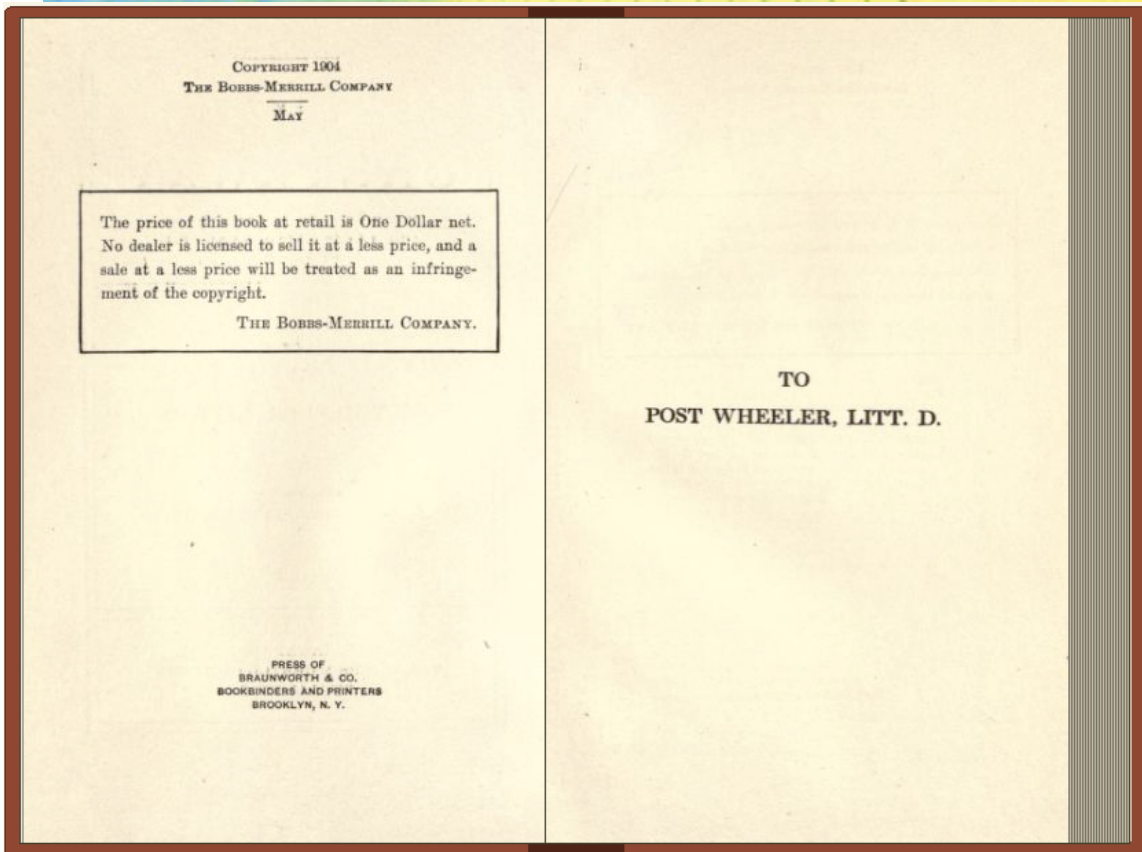


Figure 1: Notice (left page) published by the Bobbs-Merrill Co. in *The Castaway* by Hallie Erminie Rives in 1904. The attempt to use the right of distribution to control resale of the book eventually led to the landmark Supreme Court decision *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339 (1908). The case codified The First Sale Doctrine. The image is a screenshot taken from the electronic version of the novel available on the Internet Archive at <http://www.archive.org/details/castaway00riveiala>.



Figure 2: A physical copy of the novel *Anima* by Dalian Hansen (digitalKu 2007). The freedoms to modify (write in the margins, dogear the pages, etc.) and resell lawfully obtained copies of the book are protected by Fair Use, First Sale, and de minimis doctrines.

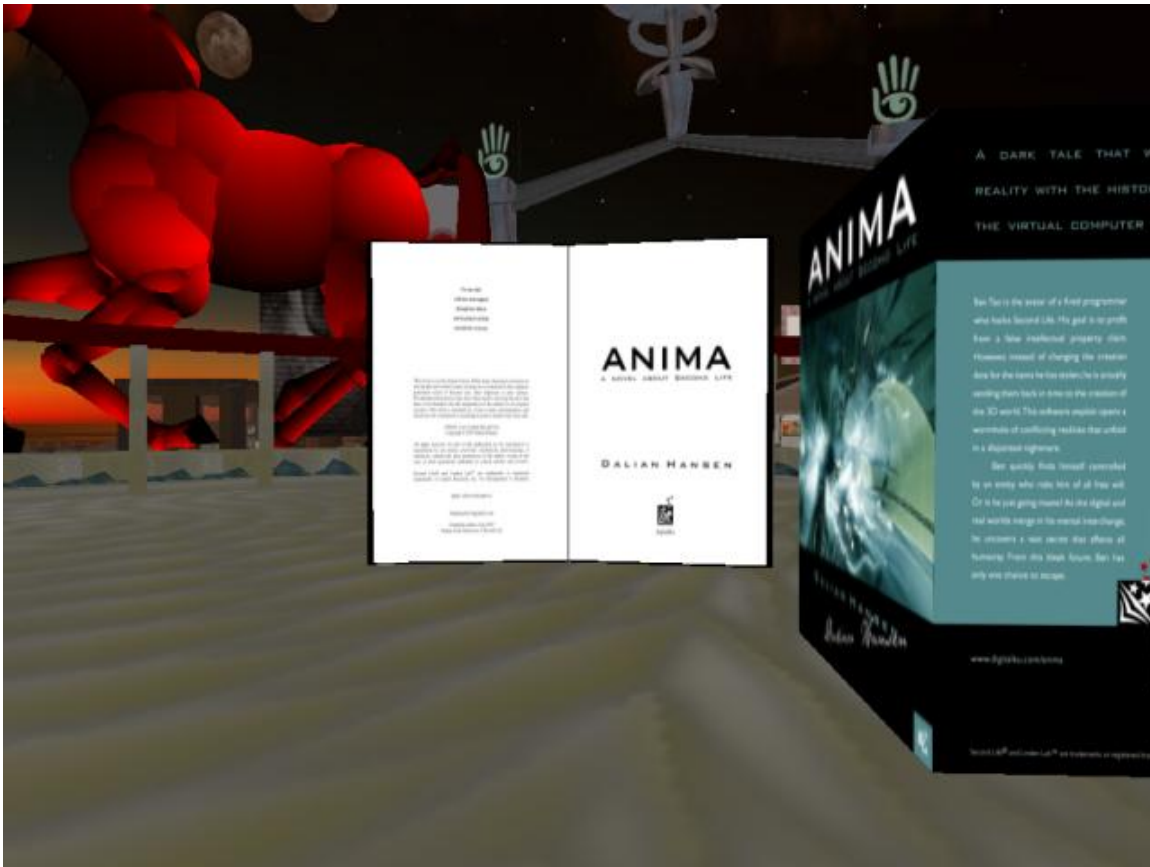


Figure 3: Virtual Copy of Anima by Dalian Hansen. This version of the novel can only be read in the virtual world of Second Life. In this case the author has disabled transfer and modification rights. The permissions system of SL partially nullifies the Fair Use, First Sale, and de minimis doctrines that operate in real life.

FULL PAPER SUBMISSION 14:
 Event-Driven Productivity Infrastructure

AUTHOR(S):
 Taylor, Hugh

ORGANIZATION(S):
 Microsoft, Inc., United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Event-Driven Productivity Infrastructure

Abstract

The world of information technology is currently experiencing the parallel emergence of two separate paradigms, Event-Driven Architecture (EDA) and Productivity Infrastructure (PI). Each of these constructs has unique portent for the ways in which people interact with data and applications. However, there is also an exciting potential for the two constructs to work together in an integrated, synergistic fashion.

EDA is an approach to Service-Oriented Architecture (SOA) that creates an “enterprise nervous system,” aware of changes in state that occur within applications, databases, as well as the outside world by publishing state information through XML to a message backbone, such as an Enterprise Service Bus (ESB). Event listeners, also connected to the ESB, distribute state change data to Service-oriented applications for processing and reaction, including human interactions. The EDA approach to enterprise architecture offers advantages in agility and segregation of concerns that benefit the utility of information systems.

Productivity infrastructure (PI) is an umbrella term to describe people’s and organization’s increasingly connected and synergistic use of phone, email, Internet, PDA, PDA, intranet, extranet, and desktop productivity applications. What was once a collection of essentially siloed productivity technologies and work flows – phone calls, emails, searching the Web, creating documents, using a PDA, and so on – are now merging into a combined infrastructure that drives personal and organization productivity. In brand name terms, productivity infrastructure is integrating the functionality of product sets such as Microsoft’s Office System, Cisco’s VOIP solutions, IBM’s Lotus suite, and Google’s Docs and gMail services, just to name a few.

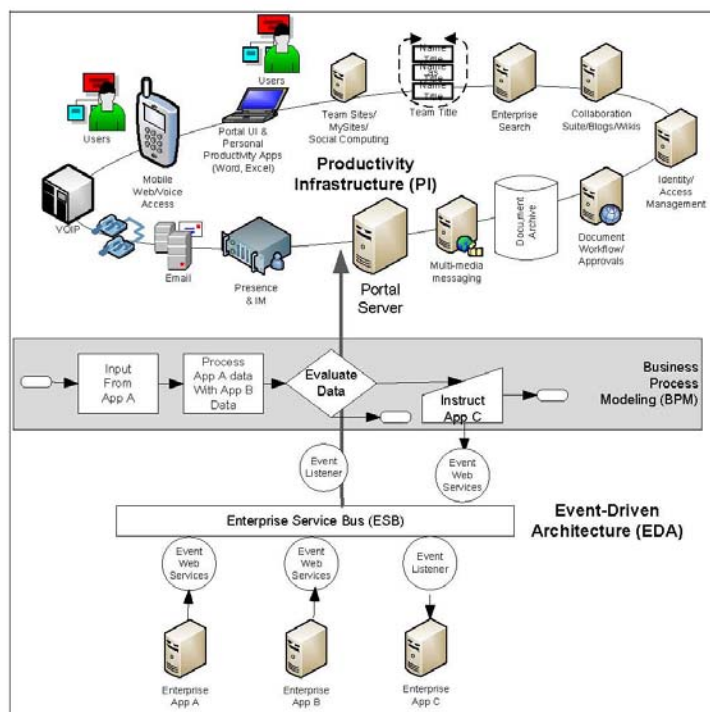
This paper explores the potential integration of EDA with productivity infrastructure. Events that originate from applications in the EDA can be published to individual human users, or collaborative groups operating in the productivity infrastructure. The integration has the potential to connect real time operations of an enterprise with Web 2.0 technologies such as blogs and wikis, as well as a range of mobile computing technologies, in addition to standard portal interfaces. The paper focuses on the ways that productivity infrastructure empowers the human thinking and decision making that is often implicit in the process flow of an EDA. It looks at scenarios where EDA-PI integration can improve the speed, quality, and cost effectiveness of end users in a large enterprise environment. Further, the paper will examine the potential for real time human activities to become events themselves, which can flow in the other direction and drive application functioning in the EDA.

Outline

1. Introduction
2. Overview of Event-Driven Architecture
 - a. The goal of EDA: enterprise “nervous system” aware of external and internal changes in state that can drive processes and reactions
 - b. How EDA works
 - i. The big picture
 - ii. Event producers
 - iii. Event listeners
 - iv. Event processors
 - v. SOAP XML and event Web services
 - vi. The role of messaging infrastructure in EDA
 - c. EDA case example
3. Productivity Infrastructure Defined
 - a. The convergence of VOIP, Web 2.0, email, portal technology, content management, IM, business intelligence, and enterprise search
 - b. Uses of PI
 - i. Individual work processing
 - ii. Unstructured task management
 - iii. Team collaboration
 - iv. Distributed group decision making in real time
 - c. PI case example
4. The Potential for EDA-PI Integration
 - a. Shortening individual and team decision cycle time
 - b. Improving quality of individual and team decision making
 - c. Improving productivity and cost effectiveness of information workers
 - d. Improving utilization of IT assets
5. Integration Scenarios
 - a. EDA publishing event data to PI
 - b. PI publishing event data to EDA through presence updates
 - c. PI publishing event data to EDA through stored enterprise search queries and RSS feeds
6. Technological Challenges to EDA-PI integration and Solutions
 - a. SIP to SOAP translation
 - b. RSS/EDA integration
 - c. Managing Security and compliance issues
7. Conclusion

*Excerpt**The Potential Benefits of EDA-PI Integration*

To see the potential for benefits of integrating EDA with PI, we should think about the advantages of linking the corporate “nervous system” of EDA with the comparable organizational nervous system of PI. With EDA, corporate systems can detect changes in state that affect business. Wherever the EDA needs human input, PI can speed up the EDA’s reaction time to the state change. PI can also improve the quality of the human input, because it can link people with data sources, and each other, with high efficiency. Ultimately, there is the potential for the creation of loops of interaction between EDA and PI, where state changes noticed by the EDA elicit reactions from people through the PI, who in turn input their own changes of state to the EDA.

**Figure X**

The integration of productivity infrastructure and EDA, connecting people and enterprise apps through an ESB and event Web services.

As shown in Figure x, the integration of productivity infrastructure and EDA can be understood by considering a simple business process model that involves inputs from two enterprise systems that must be evaluated by people. Event Web services on applications A and B publish data about their state to the ESB, and on to an application built using a Business Process Modeling (BPM) tool. The process model calls for people to assess the data presented by the states of applications A and B, and for them to make a decision about what the states mean, and then take action either by instructing application C or terminating the process without taking action. The PI is designed to notify the decision makers of the change in state. Once the decision is made, the reaction to the change in state flows back to the EDA through an event Web service located in the PI.

If there were just one person who could make the decision called for in the process, off the top of his or her head, there wouldn't be much need for the kind of elaborate setup called for in Figure X. However, let's suppose that the decision being made in the process flow is complex, high risk, multi-stakeholder, and time sensitive. Imagine, for example, that it involves the decision to manufacture goods with costly inventory. The decision could have impact on financial statements, factory capacity, even labor unions. In that kind of situation, a tight integration between the decision makers and the EDA could have a real impact on the business.

In the case of a manufacturer setting inventory levels, the time required to make a decision – the right decision, that is – is highly relevant to business success. If the manufacturer guesses wrong, and either overstocks an item whose product life is on the decline, or under stocks a hot seller, the financial results will be less than optimal. In these kind of situations, even the difference of an hour or two, or the lack of a few critical nuggets of business data, can have an impact on the bottom line. Imagine, for instance, if you decided to order a truckload of component parts for the manufacture of a product that was later determined to be unneeded. It might not be the end of the world, but it would create an accounting and logistical hassle to return the order. Multiply this type of problem across a large, global company, and the effect on earnings could reach into millions of dollars of direct and indirect costs. Consider, for example, the necessity of engaging accounting staff unnecessarily due to faulty decision making. This

In addition to offering a shorter decision cycle time, the integrated EDA-PI approach has the potential to enable a higher quality of decision than the current state of integration between PI and enterprise systems. Keeping in mind with this example that we are dealing here with decisions that cannot be automated through rules engines, consider the factors that affect the quality of decision making amongst multiple stakeholders. In our view, the quality of a decision depends on the financial consequences of the decision. The decision that saves or makes the most money is the best one. Of course, there is a range of quality decisions between best and worst, but the goal should be to strive for the best decision in the largest number of cases. This concept is known as the “decision yield”.¹

As anyone who has worked in a large, distributed organization could tell you, the quality of a decision depends on multiple interdependent factors, including knowledge of who the stakeholders are for a particular decision, equal simultaneous access to information, and equal understanding of information. Quality of decision making also depends on a productive engagement of stakeholders inside an organizational hierarchy. The higher level stakeholder may have the ability to overrule the correct decision through innate power, and the smarter subordinate may not have, or want, the opportunity to oppose the incorrect decision. Of course, productivity infrastructure cannot help an organization overcome this hierarchical flaw in process on its own. However, by providing open access to shared opinions and corporate knowledge, and real time access to multiple points of view, the hierarchy effect can be mitigated in favor of discussion and group learning.

¹ Taylor, James and Raden, Neil – *Smart Enough Systems* – Prentice Hall 2007

Event-Driven Productivity Infrastructure

Hugh Taylor

- Lecturer, UC Berkeley School of Information
- Senior Marketing Manager, Microsoft

hughta@sims.berkeley.edu

www.hughtaylorite.com

(206) 384-1241

Abstract

The world of information technology is currently experiencing the parallel emergence of two separate paradigms, Event-Driven Architecture (EDA) and Productivity Infrastructure (PI). Each of these constructs has unique portent for the ways in which people interact with data and applications. However, there is also an exciting potential for the two constructs to work together in an integrated, synergistic fashion.

EDA is an approach to Service-Oriented Architecture (SOA) that creates an “enterprise nervous system,” aware of changes in state that occur within applications, databases, as well as the outside world by publishing state information through XML to a message backbone, such as an Enterprise Service Bus (ESB). Event listeners, also connected to the ESB, distribute state change data to Service-oriented applications for processing and reaction, including human interactions. The EDA approach to enterprise architecture offers advantages in agility and segregation of concerns that benefit the utility of information systems.

Productivity infrastructure (PI) is an umbrella term to describe people’s and organization’s increasingly connected and synergistic use of phone, email, Internet, PDA, intranet, extranet, and desktop productivity applications. What was once a collection of essentially siloed productivity technologies and work flows – phone calls, emails, searching the Web, creating documents, using a PDA, and so on – are now merging into a combined infrastructure that drives personal and organization productivity. In brand name terms, productivity infrastructure is integrating the functionality of product sets such as Microsoft’s Office System, Cisco’s VOIP solutions, IBM’s Lotus suite, and Google’s Docs and gMail services, just to name a few.

This paper explores the potential integration of EDA with productivity infrastructure. Events that originate from applications in the EDA can be published to individual human users, or collaborative groups operating in the productivity infrastructure. The integration has the potential to connect real time operations of an enterprise with Web 2.0 technologies such as blogs and wikis, as well as a range of mobile computing technologies, in addition to standard portal interfaces. The paper focuses on the ways that productivity infrastructure empowers the human thinking and decision making that is often implicit in the process flow of an EDA. It looks at scenarios where EDA-PI integration can improve the speed, quality, and cost effectiveness of end users in a large enterprise environment. Further, the paper will examine the potential for real time human activities to become events themselves, which can flow in the other direction and drive application functioning in the EDA.

Table of Contents

Abstract.....	1
Introduction	3
Background of Research.....	3
EDA: A Working Systemic Definition	4
Defining an "Event"	4
EDA Overview	5
Event Producers or Publishers.....	5
Event Listeners, or Consumers	6
Event Processors	6
Event Reactions	6
Messaging Backbone	7
Assembling the Paradigmatic EDA	8
The Flavors of EDA.....	8
The Often Inadequate Human Link in the EDA	9
Overview of Productivity Infrastructure.....	10
The Potential Benefits of EDA-PI Integration.....	13
ProdCo, An EDA-PI Integration Scenario.....	16
Sales Proposal and Order Fulfillment at ProdCo	16
Prod Co's Current Productivity Tools	18
ProdCo's Proposed EDA.....	19
A Better Proposal: EDA/PI Integration	22
EDA/PI Integration Requirements	24
PIEDA's Target Architecture	26
Implementation.....	29
Conclusion	30

Introduction

Jerry Seinfeld does a routine where he wonders how mankind discovered that glue can be made from horses. He describes a scene where someone is trying unsuccessfully to stick two pieces of paper together, when, suddenly a horse rides by, causing an instant revelation of potential... "Hey, wait a minute!" he imagines the brilliant inventor saying. "How come I never thought of that before?" So it is, too, in IT. Sometimes, trends of technological innovation exist in parallel for a period time before someone realizes that they can be put to work together, for a greater effect than either one on its own. There are some great examples of this from the history of technology, including the merging of recording technology and the telephone (creating the answering machine), the joining of the QWERTY keyboard with the cathode ray tube and the CPU (the modern computer), or the phone with the computer (networked computing).

Today, we are witnessing the parallel maturing of Event-Driven Architecture (EDA), typically a subset of Service-Oriented Architecture (SOA) and productivity infrastructure, two separate, but potentially synergetic information technologies. Each is powerful in its own right, but together, they can create transcendent event driven information processing environments. This paper will explore the potential integration of EDA with productivity infrastructure. In particular, we will focus on the ways that productivity infrastructure empowers the human thinking and decision making that is often implicit in the process flow of an EDA.

To illustrate the potential for the integration of productivity infrastructure and EDA, we will use a hypothetical example of a business that connects its purchasing process and related transactional systems with its productivity infrastructure, including email, VOIP, blogs, and instant messaging. I will use the case study of a hypothetical business to understand the EDA-PI connection because the current state of the technology does not reflect its future potential. At the conclusion of this paper, I will take a quick look at several of the emerging products that are on track to fulfill the potential of EDA-PI integration.

Background of Research

The subject of connecting back-end transactional systems with messaging infrastructure is not new. There are numerous examples of comparable structures, such as Interactive Voice Response (IVR) or SNMP alerts forwarded from system management consoles to pagers, and so forth. The concept of linking instant messaging, presence, and email with XML messages used in business transactions arises frequently in standards literature, commercial software development papers, and academic discussions of messaging and Service-Oriented Architecture. For example, there is the SOAP MailTo Command in SOA.¹ Due to the inherent commercial promise of EDA-PI integration, and the commercial basis for much of the installed base of infrastructure, much of the dialogue about EDA-PI integration is occurring in the commercial sphere. For example, IBM and Nortel announced in 2007 a plan to offer an integration between SOA and VOIP.² Or, the new "Oslo" SOA offerings from Microsoft contain a foundation for linking business process modeling, Web services, and email messaging. This paper tries to go one step further and explore a vision of EDA-PI integration that has not yet been achieved in the commercial world, though one which may become a reality in the near future.

¹ <http://cpan.uwinnipeg.ca/htdocs/SOAP-Lite/MAILTO.pm.html>

² [Nortel and IBM Use SOA to Streamline Communications Among Customers, Employees, and Partners](#)

EDA: A Working Systemic Definition

Event-Driven Architecture is an approach to enterprise software architecture that is characterized by the ability of the enterprise to detect events and react to them. Brenda Michelson, a technology analyst, writes, "In an event-driven architecture, a notable thing happens inside or outside your business, which disseminates immediately to all interested parties (human or automated). The interested parties evaluate the event, and optionally take action."³ To keep the discussion simple, I am going to make the following assumption: Event-Driven Architecture in the modern sense is based on the set of XML messaging standards and inter-operability technologies collectively known as Service-Oriented Architecture (SOA). These technologies include SOAP Web Services, UDDI, WSDL, Enterprise Service Bus (ESB), and industry/commercial solutions such as Java Connector Architecture (JCA), Windows Communication Foundation (WCF), Java Message Service (JMS), and so forth. Though there are many different, equally valid approaches to the realization of EDA, I will focus on the SOA version because it now the de fact mode of implementation of the EDA paradigm.

Defining an "Event"

First, let's define what we mean by an "event." In life, an event is something that happens: a car drives by, a ball flies through your window, someone falls asleep – an action occurs. Alternatively, for our purposes, an event can also be an expected action that does not occur. If the temperature does not go down at night, that could be an "event." In systemic terms, an event generally refers to a change in "state". A change in state typically means that a data value has changed.

For example, if you make a credit card charge, there are several values that change in state as the transaction is processed. The value of your account balance will change as you make the charge, as will the state of your card from "quiet" to "transaction occurring," and so on. The shift from "balance = \$500" to "balance = \$550" is an event.

An event has three levels of detail. At one level, there is the basic fact that an event has occurred. An event has either occurred or not. This takes us to the second level of detail, which is the event definition. In order to recognize an event, an EDA must have a definition of what the event is. In the credit card example, the EDA must work with a definition of an event that says, in effect, a "change in balance" event has occurred whenever the balance value goes up. What if the balance goes down? Is that also an event? It could be, but the definition would need to take that into account. Thirdly, there is the detail of the specific event. How much has the card balance changed? In our example, the two-level event information structure would say, "So-and-so's Balance has changed. Amount = \$50." All three factors – event notification, event definition, and event detail, are necessary when designing an EDA.

For an event to occur in an EDA, it must be in digital form, or a form that computer can understand. However, that does not mean that an EDA is exclusively the preserve of digital information. A change in state can also result from non digital information being translated into digital form. For example, a digital thermometer typically has some kind of analogue temperature sensor which inputs temperature information into the sensor and results in a digital value equated to temperature. The edges of an EDA may be full of analog information that is translated into digital data in order to trigger events.

³ Michelson, Brenda – [Elemental Links 2007](#)

A key learning point here is to understand that virtually anything can be an event or trigger an event. Rainfall in Chad? Could be an event, if it is quantified and made available as a source of data. Stock market activity in Tokyo? Absolutely an event, assuming you know why you are interested. And on and on. An ideal EDA can be easily adapted to recognize events that occur anywhere. The trickiest word in the last sentence was “easily,” a simple idea that can generate a lot of discussion and complexity.

EDA Overview

Armed with a sense of what an “event” can be, we can now add some flesh to the basic EDA definition articulated by Brenda Michelson. If an event is any “notable thing” that happens inside or outside our businesses, then an EDA is the complete array of architectural elements, including design, planning, technology, organization, and so on, which enables the ability to “disseminate” the event immediately to all interested parties, human or automated. The EDA also provides the basis for interested parties to “evaluate the event, and optionally take action.”

The reason that EDA is a challenging concept is that it is so potentially broad. Just as almost any piece of data, analog or digital, can be an event, and any system in the universe can potentially be part of your EDA, where do you begin to draw the boundaries and definition of an EDA that makes sense to your organization? Though there is no bullet proof way to answer the question, I think that it makes sense to identify the main ingredients of an EDA, and build the definition from these constructs.

Event Producers or Publishers

In order to have an EDA, we must first have events... That may seem obvious, but a lot of otherwise sophisticated discussions of EDA either neglect or muddle up this central enabling concept. The EDA cannot work unless it has the ability to perceive that an event has occurred. For that to happen, the event must be created, and then published, for the components of the EDA called listeners to “hear” them.

The technologies that do this are known as “Event Publishers” or “Event Producers.” With the broad definition of an “event,” event publishers can take many different forms. Most are software programs, though an event publisher can also be a dedicated piece of hardware that translates analog data into digital form and feeds it into software that can detect an event. Here are some core ideas to keep in mind about event publishers:

Event publishers can be anywhere. Because events can occur outside of your enterprise, event publishers that relate to your EDA can be pretty much any place. Imagine the relationship between an airline EDA and the FAA radar tower. The radar tower, which serves many purposes, one of which is to be an event publisher, is completely separate from the airline’s systems, yet it is part of the EDA.

Event publishers may or may not originate the data that is contained in the event itself. In their purest form, an event publisher generates a piece of data that is formatted to be “heard” as an event in accordance with the EDA’s setup for this process. For example, a credit card processing system should automatically generate data that is EDA ready – it contains the card holder’s identifiers, the time of the transaction, the amount, the merchant name, and so on. Of course, the data was created for the purpose of charging the card, not feeding the EDA, but it serves that purpose quite readily. In contrast, other event publishers need to translate data into a format that constitutes an event according to what the EDA requires. For example, there is no inherent event pattern in the Tokyo stock exchange index unless you specifically instruct an event publisher to transmit data about the index in a manner that makes sense to your EDA’s purposes.

Event Listeners, or Consumers

Like event publishers, event listeners can be anywhere. In theory, the event listener (or consumer) has a communication link with the event publishers. That is not always the case, but we will work under that assumption for now. The event listener is a piece of technology – typically software based, but also hardware – that “knows” how to differentiate an event, as it is defined, from other data it receives.

In the simplest form of EDA, the event listener can only receive the specific event data that it is meant to hear. For example, an EDA for building security may be based on a burglar alarm whose “Event Listeners” can only hear one kind of event - the kind created by window break, detecting “event producers.” There is just one kind of event that can occur: a break in. The real world, of course is more complex, and as we progress, we will get into more involved EDA setups.

Event listeners also need to know what they are “listening for”... An application that reads the data stream of the Tokyo stock market average is not an “event listener” until it has been instructed to “listen” for some specific type of event. For instance, the event listener must know that a gain of 5% or more in the average is an “event.” A 4% gain is not an event. The event listener must be able to detect the event, and be capable of learning what the event is.

Event Processors

After an event has occurred, and been “published,” and “consumed” or “listened to” – it must be processed. It does little good to have an event that is perceived, but not handled. An EDA without Event Processors brings to mind the great joke about economists. A man gets lost in a hot air balloon. After drifting for hours, he finally sees an economist standing on the ground. The man yells out, “Where am I?” To which the economist replies, “You’re up in the air.” Unlike the economist, who may know the facts but be unable to make any sense of them, an EDA should have the capacity to interpret the events it hears.

An event processor is invariably a piece of software. While it may or may not be part of some larger, more comprehensive suite of applications, an event processor is distinctive because it has the ability to assess events, determine their importance, and generate a reaction of some kind, even if the reaction is “do nothing.”

Event Reactions

Following our chain of activities, we have an event, which is published by an event publisher, heard by an event listener, and processed by an event processor. Then, something (or nothing) needs to happen. Because “something happening” is inherently more interesting than “nothing happening,” let’s look at event reactions that require action.

Reactions to an event in an EDA vary widely, from automated application responses, to automated notifications sent either to applications or people, to purely human reactions based on business processes that occur outside of the EDA itself. In the purely automated application response category, we might see an EDA that reacts to an event by initiating an application level process. For example, in the credit card fraud example, the EDA might modify a variable value from “Normal” to “Warning” based on an event that suggested that fraud were occurring. If this reaction were coded into the event reaction, it will happen without any human involvement. Following from this, another related branch of processes might handle new charge requests on the

account differently based on a “Warning” value than from a “Normal” state. Event processors and reactions can be linked and interdependent.

The event reaction might be machine-to-human. Continuing with our example, imagine that there is a customer service representative who sees all the new “Warning” value changes, and is prompted to call the cardholder to inquire about the status of their card. This approach to EDA is dependent on the human reaction to an event, a situation that may be good or bad, depending on the desired outcome. For example, many intrusion detection systems that monitor networks for unauthorized access attempts generate a great deal of false positives. The system itself, which is essentially a tightly defined EDA, will only be as good as the person who monitors it. Indeed, there are some intrusion detections that are not monitored at all. What these defectively implemented systems do is generate logs consisting of thousands of possible intrusion records, which are essentially useless. The takeaway here is that the human reaction is very much part of the EDA design, even though it does not involve technology per se.

Finally, there are EDAs where the event reaction is based on an entirely human set of processes, following the detection and processing of the original event. For example, an EDA might generate a number of alerts a number of stock market indicators. An investor then reviews the alerts and decides whether to buy or sell. Once again, the EDA design must take into account the human reaction as part of its eventual success or failure. And, with this type of complex human decision-making, it can be quite challenging to determine what the “right” reaction should be. Experts may differ on how to react to identical sets of event alerts. As they say on Wall Street, these differences of opinion make for a good horse race...

Messaging Backbone

The fifth and final core component of an EDA is the messaging backbone, the communication infrastructure – inclusive of hardware, software, network protocols, and message formats – that enables each piece of the EDA to communicate with one another. In order to serve an EDA effectively, a messaging backbone must have several characteristics. An EDA messaging backbone needs to be as near to universal as possible, meaning that it should enable messaging across multiple network transport protocols and data formats. In other words, it should be standards based, or have the ability to mediate across multiple messaging standards. It should be pervasive, i.e. far-reaching and universally accessible. In reality, this means that it is based on the Internet, though it need not be. It should be inexpensive to develop, maintain, and modify – a lot easier said than done, but this is a critical. Cost is the “invisible hand” that has killed many great EDA initiatives. Finally, it should enable a high level of de-coupling between event producers and event consumers. In reality, this usually translates into a “publish/subscribe” or “pub/sub” set up.

The messaging backbone is arguably the most essential piece of the EDA puzzle, for without it, there can be no EDA. Without the ability to communicate, the event listeners, producers, processors, or reaction processes, cannot work. Now, you might be thinking, yes of course they can – one can always create communication interfaces between systems. Of course, this has been true for many years. The reality, though, is that proprietary interfaces, which have been the traditional way to achieve connections, are costly to develop, maintain, and modify. So costly, in fact, that they have rendered the concept of a dynamic EDA virtually impossible to realize. As I have noted earlier, I will work with the assumption that the messaging backbone of the EDA will be an Enterprise Service Bus (ESB) or equivalent.

Assembling the Paradigmatic EDA

To get to the paradigmatic EDA – the one I will use as the reference point for the rest of this paper – involves connecting event producers, event listeners, event processors, and event reactions using a common messaging backbone. Figure 1 shows what this looks like in the plainest terms. Obviously, things can get wildly more complex in real life, but I thought it would be best to start with a very stripped down paradigm example to start.

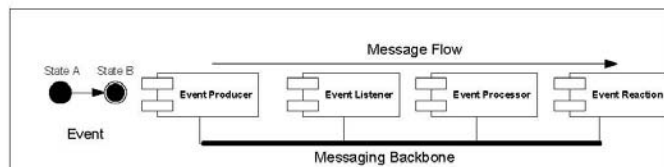


Figure 1

The paradigmatic EDA – consisting of event producer, event listener, event processor, and event reaction, all connected through a common messaging backbone

Like an “enterprise nervous system,” the EDA works as a whole, taking in signal inputs in the form of event data, processing it, and reacting to it based on some kind of intelligent model. Like the nerve endings that tell us when we are hot or cold, being touched, or falling through the air, event listeners in the EDA quantify information from the world and order it into a form that the EDA can understand. The messaging backbone of the EDA is like the nerve cells themselves, transmitting signals back and forth from various places in the body and toward the brain. The brain is like the EDA’s event processing components. It takes in event data and decides how to respond to it. The event reaction components are like our limbs. Based on the input, we take action, or not.

The Flavors of EDA

There’s an old joke about Las Vegas, where a guy wants to be seated at a restaurant and is asked, “Which section would you prefer, smoking or chain smoking?” EDA is the same. Almost every system has some degree of event orientation in it. After all, computing is essentially a matter of input, processing, and output, much like an EDA. Following from this, discussions of what an EDA actually looks like can be confusing because they tend to get overly broad and encompass, well, just about everything the IT universe. To keep our focus, let’s look at several prime models of EDA that are in use today.

According to Manas Deb, there are two essential types of EDA: Explicit and Implicit.⁴ In an explicit EDA, the event publishers send event data to known event listeners, perhaps even by direct hard coding of the event listener destination right into the event producer. Most current implementations of EDA are either partly or wholly explicit. As such they tend to rely on tight coupling of event producers and event listeners.

In contrast, an implicit EDA does not specify any dedicated connections between event producers and event listeners. The event listeners, or the event processors even, may determine the events to which they want to listen. The coupling between the event producers and listeners will be loose, or even completely de-coupled in an implicit EDA. As you might imagine, an implicit EDA is more flexible and dynamic than an explicit EDA. Historically, they have proved

⁴ Deb, Manas [EDA and Event Processing Essentials – Visual Studio Magazine 2006](#)

too complex to implement. This is beginning to change with the advent and adoption of open standards.

Within EDAs themselves, there are three basic patterns: Simple Event Processing, Event Stream Processing, and Complex Event Processing, which is known as CEP. In simple event processing, the EDA is quite narrow and simple. The event producer's function is to generate event data and send it to the event listener, which in turn processes it in whatever manner is required. The thermostat example is emblematic of a simple event processing design. The furnace gets the signal and switches on or off, and that's about it.

Event stream processing involves event processors receiving a number of signals from event producers (via the event listeners) but only reacting when certain criteria are met. For example, the thermometer may send the temperature data every two seconds, but the event processor (thermostat) ignores all but the relevant "switch on" data point, which activates the furnace only when it is observed.

Complex event processing (CEP) takes the EDA to another level, but enabling it to react to multiple events under multiple logical conditions. So, for the sake of argument, let's say that we only want to buy US Treasury bonds when the Nikkei hits a certain number, and unemployment figures dip below 5%. The CEP listens to multiple event streams, and knows how to correlate them in a logical manner according to the objectives of the EDA. In other words, the CEP starts to "think" like a person, taking in data from unrelated sources on the fly, differentiating between useful and useless information, and acting accordingly. This is the ideal of the complex, implicit, dynamic EDA.

The Often Inadequate Human Link in the EDA

There are many instances where the corporate "nervous system" of EDA loops through a human decision making process. In the airline traffic case, it was the flight operations managers who were called upon to make critical decisions about flight prioritization based on input from the EDA. In the anti-money laundering case, bank fraud staff were fed information about suspicious transactions for their review and decision on actions. This EDA-human connection makes sense much of the time. Indeed, there is often no substitute for a person, and his or her awareness of multiple influencing factors, in a business decision making process. Artificial intelligence is suitable in some cases to make or support human decisions, but even in cases where decision making can be automated, there is frequently the need for a person to take responsibility for the decision. Alas, there is still no way to hold a computer accountable for the consequences of a decision that causes an airliner to crash or money to be stolen from a bank account. Given the inevitable presence of people in EDA-based decision process flows, one of the big challenges is relative inefficiency of human decision making.

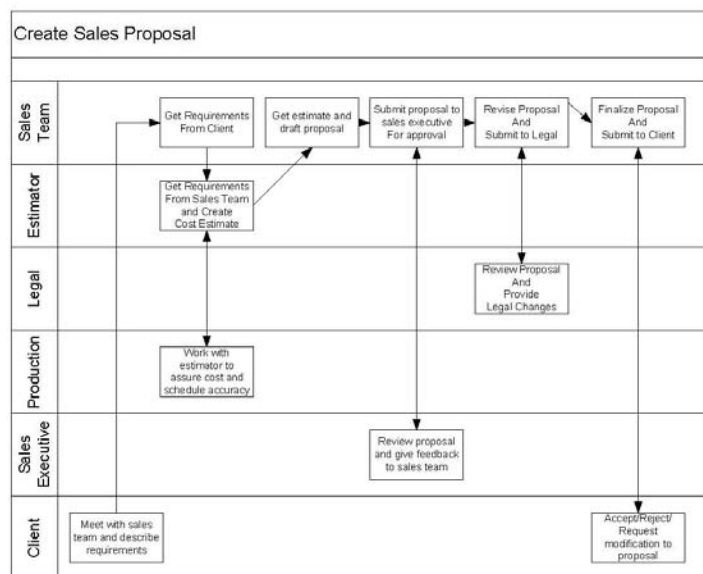
Unlike computers, people are extraordinarily inefficient at decision making and are, in fact, quite high maintenance. While a computer can execute a decision algorithm at any time within a fraction of a second, people need to be present (awake and focused) to make a specific decision at a specific time. This is not efficient, and in some cases, may be harmful to the business process that the EDA is meant to serve. And, in many cases, people need to make decisions in groups, a situation where the inefficiency of communication compounds the delay and quality of the decision. For example, if the air traffic EDA enables rapid decisions about prioritizing flight departure times, but the key decision maker needs to consult with a superior, who is out to lunch, the whole process could be delayed to the point where it compromises the whole intent of the system.

In other cases, people may need to access external sources of information in order to make decisions that feed into an EDA decision making process. The bank anti-fraud staffer may have to review scanned document images to compare signatures on old checks before making a judgment call about whether to escalate a fraud investigation or notify law enforcement. This type of manual, or semi-manual process, can cause harmful inefficiency for the EDA. In this instance, the anti-fraud staffer might have to manually write down the name of the suspected account, exit the EDA interface, open a records management application and conduct a search for matching documents. After manually reviewing the documents, he or she may have to share the findings with other anti-fraud staffers or document review specialists before making a decision to escalate the case. Such a discussion could involve a combination of email, phone, fax, or instant messaging. None of this is particularly horrible, but the cumulative effect of faults in the communication and manual process flows – spread out across multiple concurrent fraud cases – can result in a significant drag on performance and sub-optimal anti-fraud efforts. The solution to these types of challenges is known broadly as “productivity infrastructure,” and it is maturing today at a rapid rate.

Overview of Productivity Infrastructure

Productivity infrastructure (PI) is an umbrella term to describe people's and organization's increasingly connected and synergistic use of phone, email, Internet, PDA, intranet, extranet, and desktop productivity applications. What was once a collection of essentially siloed productivity technologies and work flows – phone calls, emails, searching the Web, creating documents, using a PDA, and so on – are now merging into a combined infrastructure that drives personal and organization productivity. In brand name terms, productivity infrastructure is integrating the functionality of product sets such as Microsoft's Office System, Cisco's VOIP solutions, IBM's Lotus suite, and Google's Docs and gMail services, just to name a few.

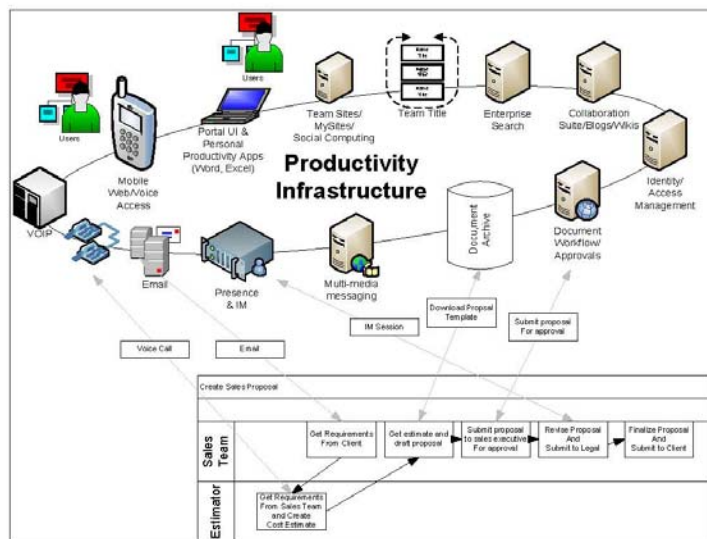
To understand the importance and impact of productivity infrastructure, let's use the creation of a sales proposal as a baseline example of the kind of unstructured type of workflow that typically challenges information workers to be productive. In contrast to structured tasks, such as those performed by customer service agents at a call center, a great deal of business work today involves unstructured tasks, which are unpredictable in terms of work flow step order, location of needed information and stakeholder identities, roles, and responsibilities. In the case of creating a sales proposal, there may be a number of different approval patterns, issues to be resolved, and decision makers involved at any given time. The processes, people, and underlying data and documents required to create the sales proposal may change from case to case. Though the process will contain the basic flow shown in figure 2, in reality each situation will be slightly different. Managing this subjectivity within a tight time frame is the essence of productivity infrastructure.

**Figure 2**

In the process flow of creating a sales proposal, multiple people and groups must collaborate and share documents and information, often in real time.

Each step in the creation of the sales proposal, as shown in Figure 3, involves multiple people, often from different work teams, in the sharing of information, documents, and knowledge. The more efficiently the people involved in completing this multi-step, multi-player process can work get their work done, the better off the organization will be in productivity terms. The impetus behind the development of productivity infrastructure is the drive to enable workers in unstructured information work to get more done in less time, with less expenditure of resources.

As Figure 3 shows, the steps in the sales proposal development process map to capabilities in the productivity infrastructure. In this example, VOIP technology speeds up the process of connecting the customer with the sales rep by automatically connecting a landline call to a mobile device. All participants in the process use email to communicate, with relevant stakeholders able to share links to document repositories where proposal templates and other data are stored for common use. A portal UI links stakeholders to the workflow management process as well as collaboration sites that contain blogs and wikis that publish up to date information needed for crafting the proposal. Social computing technologies allows individuals to understand connections between stakeholders that may not be apparent through job titles or task assignments. Team sites enable stakeholders to make sure that their proposal is in alignment with team objectives and business goals, or clarify approvals required for the proposal. The capability to see presence enables real time instant messaging or web conferencing that facilitates rapid resolution of open issues for the proposal.

**Figure 3**

Productivity infrastructure, which links workflow, collaboration, email, phone, mobile, document repositories, and Web 2.0 technologies such as blogs, wikis, mysites, and social computing, can drive efficiencies in the unstructured, multi-player, iterative process flow required to create a sales proposal.

If everyone involved in creating the sales proposal were sitting in the same room at all times, then there would be no reason for the investment in productivity infrastructure. In fact, that's how business worked until about 1890, when the telephone started to move people away from physical proximity to their business partners. Today, of course, the groups of people involved in getting business processes accomplished are almost never all together, and certainly not in any reliable pattern or schedule that will allow them to get time sensitive tasks done. And today, as we often find, it's not just being able to communicate with others in a collaborative process that makes things flow smoothly. To get the tasks accomplished, each participant needs to know who the other players are, who they report to, what their priorities are. On top of all that, participants also often need access to information that is not easily located without the assistance of others in the group. To make access to needed information available without requiring time consuming conversations or emails, participants need to be able to search for and find what they are looking for on their own.

Productivity infrastructure has the potential to drive more efficient work flows for unstructured tasks, assuming it is implemented properly. The time cycle for completing the entire process flow for creating a sales proposal, for example, becomes shorter with good productivity infrastructure, and the time investment of each participant goes down as well. Ideally, the accuracy and quality of the final product improves as an added bonus. However, productivity infrastructure can be complex to deploy, as it can raise some odd security and compliance concerns. I mention this here just to assure you that I am not all starry-eyed about the ease of deploying such a comprehensive and interconnected infrastructure, and neither should you. Nevertheless, I do believe that productivity infrastructure, once in place, has portent for EDA.

The Potential Benefits of EDA-PI Integration

To see the potential for benefits of integrating EDA with PI, we should think about the advantages of linking the corporate “nervous system” of EDA with the comparable organizational nervous system of PI. With EDA, corporate systems can detect changes in state that affect business. Wherever the EDA needs human input, PI can speed up the EDA’s reaction time to the state change. PI can also improve the quality of the human input, because it can link people with data sources, and each other, with high efficiency. Ultimately, there is the potential for the creation of loops of interaction between EDA and PI, where state changes noticed by the EDA elicit reactions from people through the PI, who in turn input their own changes of state to the EDA.

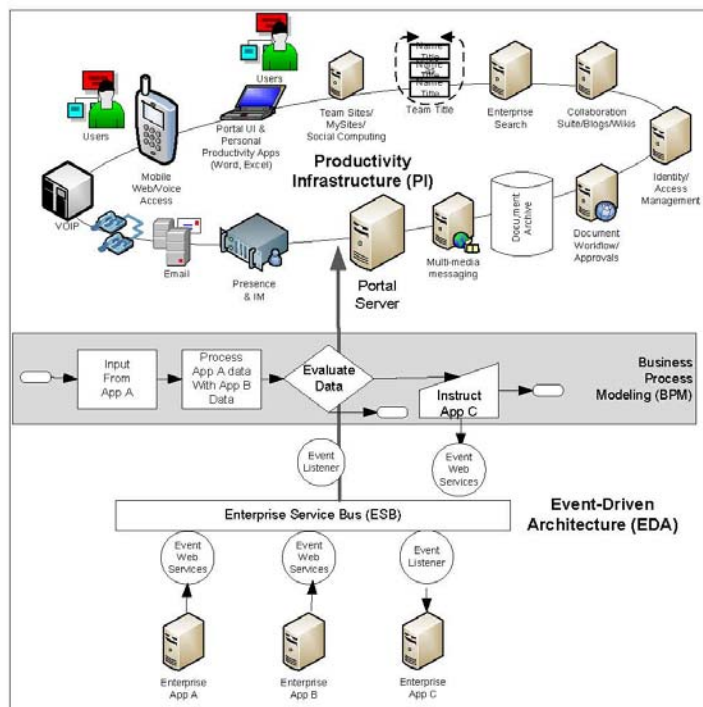


Figure 4

The integration of productivity infrastructure and EDA, connecting people and enterprise apps through an ESB and event Web services.

As shown in Figure 4, the integration of productivity infrastructure and EDA can be understood by considering a simple business process model that involves inputs from two enterprise systems that must be evaluated by people. Event Web services on applications A and B publish data about their state to the ESB, and on to an application built using a Business Process Modeling (BPM) tool. The process model calls for people to assess the data presented by

the states of applications and B, and for them to make a decision about what the states mean, and then take action either by instructing application C or terminating the process without taking action. The PI is designed to notify the decision makers of the change in state. Once the decision is made, the reaction to the change in state flows back to the EDA through an event Web service located in the PI.

If there were just one person who could make the decision called for in the process, off the top of his or her head, there wouldn't be much need for the kind of elaborate setup called for in Figure 4. However, let's suppose that the decision being made in the process flow is complex, high risk, multi-stakeholder, and time sensitive. Imagine, for example, that it involves the decision to manufacture goods with costly inventory. The decision could have impact on financial statements, factory capacity, even labor unions. In that kind of situation, a tight integration between the decision makers and the EDA could have a real impact on the business.

In the case of a manufacturer setting inventory levels, the time required to make a decision – the right decision, that is – is highly relevant to business success. If the manufacturer guesses wrong, and either overstocks an item whose product life is on the decline, or under stocks a hot seller, the financial results will be less than optimal. In these kind of situations, even the difference of an hour or two, or the lack of a few critical nuggets of business data, can have an impact on the bottom line. Imagine, for instance, if you decided to order a truckload of component parts for the manufacture of a product that was later determined to be unneeded. It might not be the end of the world, but it would create an accounting and logistical hassle to return the order. Multiply this type of problem across a large, global company, and the effect on earnings could reach into millions of dollars of direct and indirect costs. Consider, for example, the necessity of engaging accounting staff unnecessarily due to faulty decision making. This ???

In addition to offering a shorter decision cycle time, the integrated EDA-PI approach has the potential to enable a higher quality of decision than the current state of integration between PI and enterprise systems. Keeping in mind with this example that we are dealing here with decisions that cannot be automated through rules engines, consider the factors that affect the quality of decision making amongst multiple stakeholders. In our view, the quality of a decision depends on the financial consequences of the decision. The decision that saves or makes the most money is the best one. Of course, there is a range of quality decisions between best and worst, but the goal should be to strive for the best decision in the largest number of cases. This concept is known as the "decision yield".⁵

As anyone who has worked in a large, distributed organization could tell you, the quality of a decision depends on multiple interdependent factors, including knowledge of who the stakeholders are for a particular decision, equal simultaneous access to information, and equal understanding of information. Quality of decision making also depends on a productive engagement of stakeholders inside an organizational hierarchy. The higher level stakeholder may have the ability to overrule the correct decision through innate power, and the smarter subordinate may not have, or want, the opportunity to oppose the incorrect decision. Of course, productivity infrastructure cannot help an organization overcome this hierarchical flaw in process on its own. However, by providing open access to shared opinions and corporate knowledge, and real time access to multiple points of view, the hierarchy effect can be mitigated in favor of discussion and group learning.

Integrating EDA and PI can improve the efficiency of the information workers who are tasked with making business decisions. Ultimately, this can result in reductions in overhead or increased utilization of staff for strategic business purposes. I felt this point was worth making because I have heard many dialogues about the value of SOA and EDA that make the assumption that there is a high efficiency analysis apparatus available to parse the output of these systems.

⁵ Taylor, James and Raden, Neil – *Smart Enough Systems* – Prentice Hall 2007

This is not necessarily so, and indeed, a lot of approaches to SOA dead end into an empty seat called “stakeholders” and fail to generate good ROI as a result.

The Productivity Infrastructure itself can function as an event producer as well. A simple example might be the presence of a stakeholder being announced as an event. Another example, which touches on an exciting new area of PI, is the concept of “active search” within the enterprise and its potential to function as an event producer from within the PI.

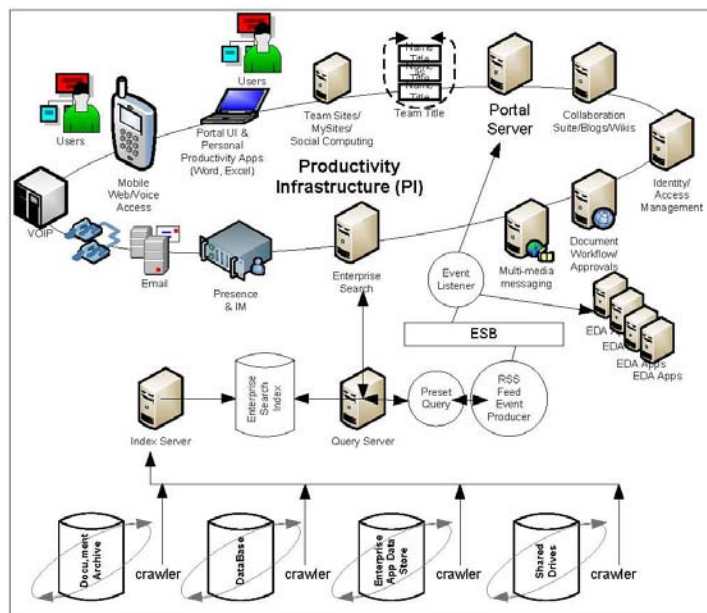


Figure 5

Active search involves the use of an enterprise search solution, which “crawls” repositories of unstructured data, indexes them, and then sends RSS feeds in response to preset search queries. The RSS feeds, which function to indicate the presence of specific information in unstructured data, can serve as event producers.

Figure 5 depicts an enterprise search solution allowing a productivity infrastructure to function as an event producer. To see how this works, we must first understand the process of an enterprise search solution, which is an increasingly common fixture in today’s enterprises. Like a Web search engine, the enterprise search solution contains three core elements: a query server, an index server, and “crawlers,” which read through documents and other data sources and feed their findings into the index server. The enterprise search index, like their corollaries in the Web search world, is a massive and exhaustive directory of information located within the enterprise. The enterprise search solution operates by taking queries from end users through a front end UI (eg a search box in a portal interface), sending the queries through the query server to the index and returning matching results back to the end user through the UI.

Some enterprise search solutions offer the ability to conduct active search, a process wherein certain queries are stored and continually re-run, with the search results being published to end users through RSS (Real Simple Syndication, a form of XML). So, for example, an end user in a real estate development firm could use an enterprise search solution to query the company's internal document libraries for a data that matches the keyword of a particular neighborhood. If the user types "Chelsea" into the search box on the portal UI, that query will return any documents or other indexed data files that contain the word "Chelsea," and the user could then learn about projects or people involving Chelsea. In an active search mode, the user could store the "Chelsea" search and instruct the search solution to issue him an RSS feed every day that contained the latest search results for "Chelsea" without the user having to run the query every day himself.

From an EDA perspective, it is possible to imagine how this active search function could turn into an event producer. The stored query, and resulting RSS feed, could be designed to publish changes in state that exist within unstructured data environments. In our real estate example, the query might feed into an algorithm that determines whether specific people or company divisions are working in Chelsea, a change in state that could drive action through the EDA. A CRM system attached to the EDA could flag the activity in Chelsea, based on the event data published through the enterprise search solution and PI, for follow up by account executives in the neighborhood.

There is even the potential for EDA and PI to inter operate as looping, connected halves of a bigger EDA. Events published out through the EDA inform stakeholders and drive action through the PI, which in turn publishes back event data about stakeholder activities, presence, and data creation through the PI. Admittedly, this level of sophistication is fairly futuristic, even for this forward looking paper. However, I believe the potential for productivity improvements and information worker empowerment through the integration of EDA and PI are powerful and promising.

ProdCo, An EDA-PI Integration Scenario

To see how EDA and PI could be integrated, I will use the example of a custom manufacturing business. In order to optimize the learning experience, we are going to keep the example fairly simple and focus on the aspects of this hypothetical business that are most general to all businesses. This company, which I'll call ProdCo, could stand in for a mass of businesses that perform custom services on a job-by-job basis. Within ProdCo, I will focus on the sales proposal and order fulfillment process to highlight the potential EDA-PI integration.

Sales Proposal and Order Fulfillment at ProdCo

The sales and fulfillment process at ProdCo is touched by a group of teams and individual stakeholders. To get a proposal out to a client and then fulfill on the order, the Sales, Marketing, Production, Legal, Finance, and C-Suite executives need to be involved. Sales, of course, is the main point of contact with the client. Marketing engages in the process to assist with pricing and discount programs that may be tied to particular campaigns. Finance is involved to assure that the pricing and costing of the job are appropriate, and that the HR aspects of the job are properly

considered. Finance also weighs in with sourcing decisions and execution. Production is responsible for actually doing the work. The C-suite has oversight, especially if the job is large or strategically significant.

Figure 6 shows the essential flow for ProdCo's proposal-fulfillment-assessment process. The sales team drafts the proposal and circulates it through reviews by legal, production, finance, and the C-suite. If there are revisions, the review cycle may repeat in whole or part. Once the client approves the proposal, the job goes into production, and is fulfilled. Finance becomes involved again for sourcing of materials and overseeing labor expenses, as well as invoicing and collection. At the end of the process flow, the finance department publishes the results of the job – if it made or lost money compared to the estimate – and all other departments receive this information and update their own knowledge bases. Or not...

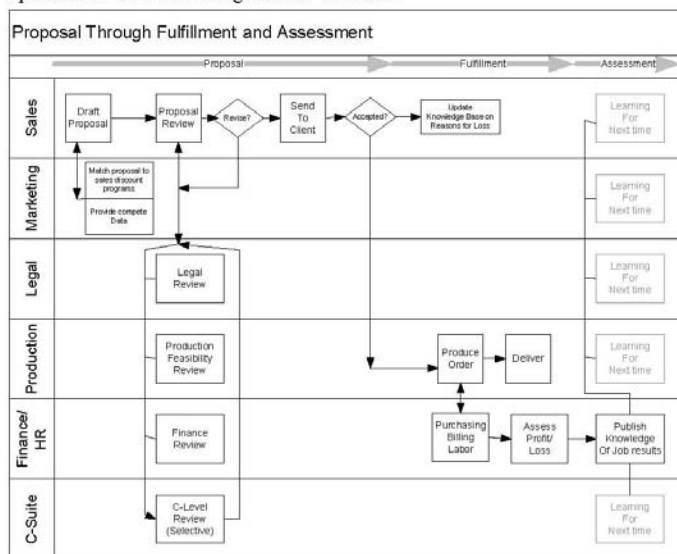


Figure 6
ProdCo's process flow for creating a sales proposal, fulfilling the job, and assessing its profitability.

There are a couple of flaws in this process design, though to be fair to ProdCo, it's about as good as it can be given the current state of technology. For one thing, the review loops for the proposal may be a lot more complicated and subjective than any process model can approximate. And, the assumption that the process flow makes is that everyone has access to relevant information on a timely basis. The marketing department may not know, for example, that certain types of projects lose money, so they ought to be dropped from the discount plan, and so forth. Most problematic, though, is that the process is very inefficient. The people and groups involved in this process waste time managing and finding the information they need to get the process finished as well as communicating with one another. A closer look at the way ProdCo has set up its productivity software and enterprise architecture can reveal some of the causes of this inefficiency.

Prod Co's Current Productivity Tools

Figure 7 shows how ProdCo has set up its productivity tools and enterprise applications. The sales and marketing teams have access to the CRM system, while finance and production use the ERP solution. Collaboration inside each team, and between teams, is a fairly ad-hoc affair, with stakeholders emailing files back and forth and saving them on departmental shared drives that each team can access through a portal interface. It is possible for a non team-member to access a portal, but that person must first be granted access rights by a departmental administrator. Each team portal has calendaring capabilities, and everyone is able to schedule meeting the email suite. The c-suite executives use a business intelligence dashboard that is fed by the finance staff because it does not tie directly into the ERP system.

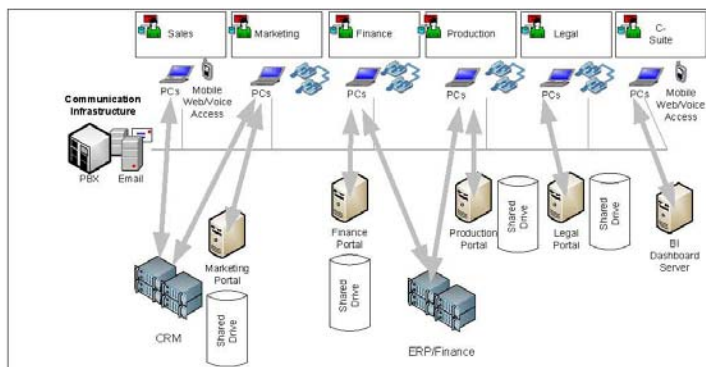


Figure 7

The ProdCo teams involved in creating sales proposals and fulfilling orders use two enterprise systems – CRM and ERP – and a slew of ad-hoc productivity solutions, including share drives, email, and voice communication.

ProdCo's productivity solutions and enterprise applications are too siloed to be highly efficient. There is no efficient way to share information or documents across the entire company – at least in a way that does not open access to the document to every single employee, which nobody wants. Connections between systems are haphazard or non-existent. So, for example, there is no way to seamlessly import the terms of a sales proposal into the ERP system. It must be re-entered when the order goes into production. Approvals on each phase of a proposal, and its subsequent production phase, are conducted by email and phone. The ERP system does have automated approval functions, where executives can sign off on purchase orders, overtime schedules, and so forth. However, the problem is that these approvers must often communicate with others, such as the legal department, before proceeding. The efficiency of the automated approval function is mitigated by the slow, unstructured process of communication within the firm. And, there is no way to keep track of recurring patterns of unstructured workflow that could save time in the future.

For example, imagine that certain types of orders require materials to be sourced from Mexico. The procurement staff has learned from experience that in the summer months, the heat inside the trucks coming from Mexico is so intense that it can ruin the parts in transit. They know now either to order these parts in advance, or actually pay for a refrigerated truck. Of course, this is more expensive, though paying the expense is preferable to delays and missing parts. However, the procurement staff has no way to keep this relevant fact in front of all stakeholders at all times.

The ERP system has a “notes” section, where the procurement staff can write down a reminder to order a refrigerated truck with that SKU. Yet, when the sales team wants to make a deal, or the marketing department wants to create a discount campaign, they do so without realizing that their margin is lower than normal on the item that includes the Mexican components. The c-suite, too, may lack visibility into the issue, and wonder why margins are low on this type of product.

ProdCo's Proposed EDA

Being wise and forward-looking, ProdCo has decided to invest in shifting its architecture towards SOA and EDA. Figure 8 shows how this would work. ProdCo would install an Enterprise Service Bus as an integration layer that exposes the functionality of both the CRM and ERP systems as sets of Web Services. A portal server would provide access to the various department portals and also make certain CRM and ERP functions available to users who didn't have access to either the CRM or ERP client. The departmental portals themselves would remain essentially untouched, though their provisioning could now be centrally controlled and extension of existing portals could be governed more thoroughly than before. Communications would remain a silo.

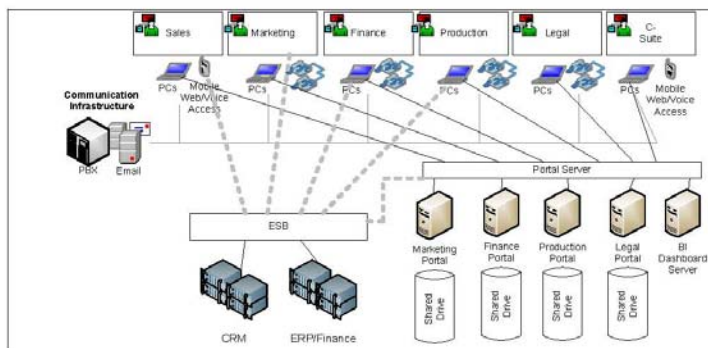
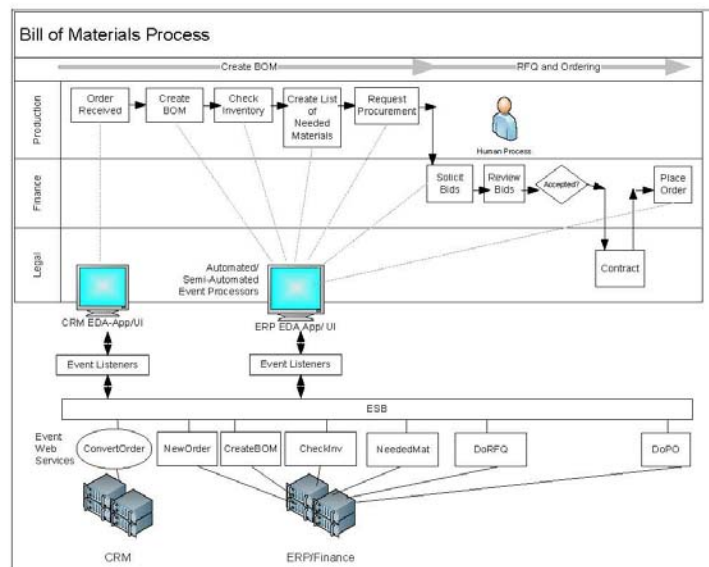


Figure 8

ProdCo is considering a move towards an SOA and EDA, where the CRM and ERP/Finance applications would be exposed as Web Services through the Enterprise Service Bus. The departmental portals would be accessed through a central portal server, though their content and administration would remain essentially siloed. The communication infrastructure would remain unchanged.

To appreciate how the EDA approach would work for ProdCo, let's look a small part of the production process, the ordering of supplies based on a bill of materials. Figure 9 shows the flow for the bill of materials (BOM) process as well as the matchup between the process steps and the Web services enabled in the EDA. Certain steps, though, such as selecting the winning bids from the RFQ, are still human processes and cannot be fully automated. Other steps, such as requesting procurement, may be semi-automated, wherein the ERP system does all the work once a person has approved the request for RFQ.

**Figure 9**

ProdCo's EDA exemplified through the process flow of creating a bill of materials and requesting quotes (RFQs) from suppliers.

As shown in Figure 9, the bill of materials process consists of several automated/semi-automated steps that work off the event Web Services, as well as a few that are completely human. Sales proposals exist in the CRM system, but when a customer makes the decision to buy, under the new EDA approach, the CRM system's ConvertOrder Web service publishes the state change of the proposal from "Pending" to "Accepted." The event listener in the ERP EDA application receives this state change information and, in reaction, takes the order data from the transfers it into the ERP system. This step could be fully automated, meaning it could occur simply because of programmatic instruction, or it could be semi-automated, where a person gets a signal through the UI that the order is ready to convert and the human action of clicking a button on the interface actually completes the transfer from CRM to ERP systems.

The process flow then follows a series of pub/sub steps. The ConvertOrder web service in the CRM system results in the creation of a new order for production in the ERP system. The NewOrder Web services publishes that its state is "New," which is listened to by the CreateBOM (Create Bill of Materials) Web Service. The Web Services CheckInv (check inventory) and NeededMat (Needed Materials) create a list of materials that are needed for the job but which are not in stock. The DoRFQ Web Service is activated by the NeededMat Web Service's publishing of its change of state (from Nothing to List). Each of these Web Services can be invoked manually or through an automated process.

Sending the RFQs out to vendors, though, is a manual step even if the actual work is done by the system. A procurement person should be active in the process to make sure that the RFQ goes to suppliers who are appropriate for the materials needed. However, even this could be a fully automated process that reacts to the event of "DoRFQ" by generating a Request For Quote

electronically and sending it to pre-approved vendors through email or online notification calling them to a vendor portal.

Assuming the EDA is implemented correctly, it should be able to confer some advantages to ProdCo's operations. It renders the back end more flexible than a conventionally architected system. The EDA makes it simpler, faster, and cheaper to implement changes in the process flow for RFQs, and enables streamlined reporting and aggregation of data for consolidation of purchase orders and vendor management. The standards-based interfaces to the end user are also potentially more cost effective to maintain and modify as changes take place in the process flow over time.

However, the human decision elements of the process flow are not much affected by the EDA. The act of soliciting bids from vendors is still either wholly or partly human, as is the selection of winning bids. It would be possible to automate both of those processes, though, and even include some fairly sophisticated rules sets to assure best practices in procurement. Taking the human beings out of the picture for selecting vendors, for example, might occur if ProdCo could implement a set of business rules that awarded contracts to vendors with the lowest price, the best ranking for quality from production operations, and a consistent track record of on time delivery. If this automation were implemented, the bill of materials process could proceed seamlessly without the messiness and delay of human actions. This might be a desirable goal, and it might even be possible. Yet, if we are realistic and look at our own experiences in corporate life, we will realize that this level of automation is not practical or wise.

Returning to our example of goods from Mexico that melt in hot trucks in the Arizona sun, (and I'm not making this up, either. It happened to a colleague of mine) we can see that there are still instances where there is a need for subjective human knowledge to get to the best possible business result from a process flow, even with an advanced EDA in place. And, this is where the efficiency of ProdCo's EDA starts to falter.

As the automated, or semi-automated process of converting an order from proposal to sale, and the derivative bill of materials/RFQ steps cascade out from the entry of the sale into the ERP system, we still dead end at a person – or group – that needs to decide which vendors get to bid on the order. In the best case scenario, the vendor selection is done by an experienced person who understands all the subjective issues involved and acts promptly and decisively. In a less rewarding scenario, the decision is made by a distributed group of people who may individually lack the knowledge of the subjective challenges to getting an optimal procurement accomplished. There are many scenarios in between, such as an inexperienced procurement person who cannot process the vendor selection quickly due to lack of information, or one who makes the wrong decisions based on lack of knowledge, or even a lack of awareness that certain types of knowledge are required to make the decision. In this situation, we see the promotion of the person who knows that one must request a refrigerated truck from Mexico, and his or her replacement being a person who doesn't have any idea that such a problem exists. The new person will proceed, with the best of intentions, to repeat a mistake that has long since been solved.

There are several aspects to this poor quality decision cycle. If the communication process that connects the people in the decision loop is detached from the EDA and procurement interfaces, then the communication itself risks being inefficient and inaccurate. For example, imagine that the vendor selection process is dependent on people reading long email threads from the bottom up, assessing the situation, and making recommendations. (Surely, we've all been there...) This communication pattern is less than ideal for rendering a consistent, rapid set of correct decisions. However, due to the subjective nature of most unstructured processes, such as selecting a vendor using group knowledge as the basis for the decision, email threading is probably unavoidable. At the very least, it is clumsy and unreliable. Even if everyone involved is paying attention and very well informed, the process could bog down if one or more

participants is unreachable (or un-findable) – to the extent that production orders could back up due to communication breakdowns in the procurement process.

ProdCo's EDA also lacks the capacity to store organizational knowledge. The specialized knowledge about the subtleties of procurement is not stored in a fashion where stakeholders can easily find or use the information. One enterprising person might create a procurement best practices document that he or she can use, and perhaps even share with others. But, if that document is lost on a shared network drive, its contents may never reach other stakeholders who need it. If the creator of the document moves on, the document will likely disappear. A dedicated and well-managed procurement team could also create an intranet site on the portal server for collection of practices and settling of decision issues. This approach has some benefits over a totally uncoordinated procurement method, but it might still result in time lags and communication mistakes if the stakeholders still need to toggle back and forth between their intranet team site, email, and the actual ERP system to make and implement decisions. At the very least, it is a largely unrepeatable pattern.

The integration of ProdCo's Productivity Infrastructure with its EDA is a major contributor to solving this dilemma, where lost knowledge, lack of knowledge, and poor communication mitigates the positive impact of an EDA and improvement business process management. PI integration cannot solve this problem all by itself, of course. There are myriad challenges related to training, knowledge preservation, best practices documentation, and so forth, that are required to assure good human decision cycles in a process flow. However, the chances of ProdCo attaining the best human decisions in the EDA environment are greatly enhanced by the integration of the EDA and the PI. Without this integration, the likelihood that ProdCo will optimize the organizational impact of the EDA are slight.

A Better Proposal: EDA/PI Integration

I may be displaying some hubris by characterizing EDA/PI integration as a better proposal. The truth is, it's a very new area and still quite theoretical, even more so than EDA, which is itself somewhat more of a vision than a reality. However, as we go through the potentialities of EDA/PI integration in the ProdCo case, I think you will see some exciting possibilities for improving the way work gets done in practical terms.

Unstructured tasks, such as procurement staffers rounding up best practices data from diverse stakeholders, tend to be messy and unpredictable. Given that reality, it is challenging to design any kind of technological solution that will consistently solve the problem and make the unstructured tasks faster, simpler, and cheaper to execute. In fact, there really isn't much in the way of standard language or practice to even describe the kind of problems that IT solutions need to solve for unstructured tasks. For this reason, I am going to attempt to work backwards from identification of unstructured task problems, to causes that can be remedied by EDA/PI integration, and build a solution approach from there.

First, what problems does ProdCo face with procurement, and which can be traced back to inefficient unstructured tasks and poor EDA integration? Wasting of time ranks high on the list because slower procurement typically translates into slower production. Then there is the loss of knowledge over time, which can result in slow procurement or other costly errors in production. Then, there are just plain mistakes made through poor communication or inadequate decision making processes.

Backing out of these problems, we can get to a set of business objectives for the integrated EDA/PI. ProdCo needs to have rapid procurement that is accurate. They need high quality and consistent knowledge transfer as team members move in and out of roles. And, they need

transparent, well-documented decision processes that result in accurate, timely procurement without imposing an undue administrative burden. This last point is relevant because it is almost always possible to cure a process by larding it up with many onerous administrative tasks and parameters. The net effect, though, is usually counter to the goal of efficiency. Getting the right process in place without choking the team members with bureaucracy is a fine balancing act.

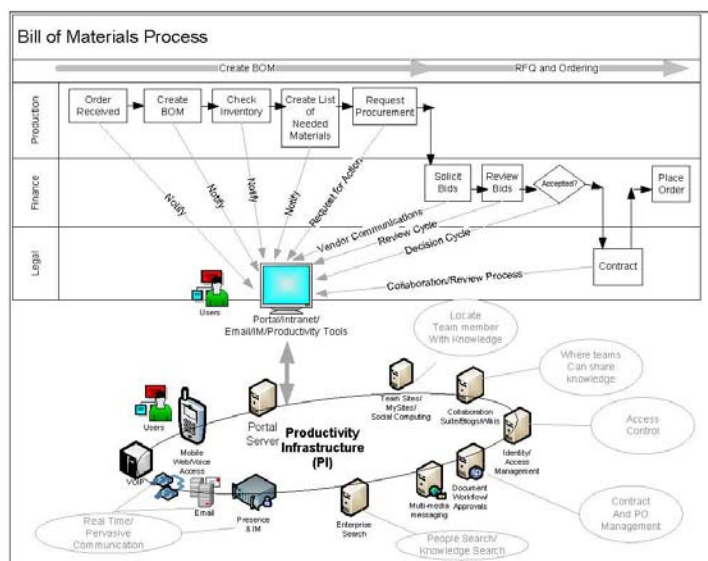


Figure 10

Matching the RFQ process flow to a productivity infrastructure, showing the connections between process steps and collaboration, communication, and knowledge sharing areas of the PI

In Figure 10, we see how productivity infrastructure can help ProdCo's procurement staff collaborate with one another, as well as other groups, share knowledge, and communicate efficiently in the fulfillment of the RFQ process. In the early steps of the process, from "Order Received" through "Create List of Needed Materials," the procurement staff are notified of changes in order status and impending RFQ workflow. The specific mechanism of notification could vary, though it would probably be an email alert or a change in an order status screen on the intranet. As the procurement staff need to solicit and review bids, they can use the team sites and search features of the PI to find expertise that may rest with individuals throughout the company. In our example, if the procurement staffer searched for the name of the product that gets shipped from Mexico, he or she might be directed to a wiki or blog that communicates the salient details of shipment in hot weather that would enable even an inexperienced procurement staffer to avoid the problem that has plagued others. Once that kind of knowledge is extant and searchable, it is harder for the organization to lose. The Web 2.0 type of features that allow individual users to create their own material easily – but also securely – is an underpinning of successful PI.

Stakeholders in the process can work on documents, such as contracts and RFQs, in a virtual collaboration environment and document management system. Throughout the process is communication, pervasive and real time, through email, phone, mobile devices, IM, and Web

meetings. The net effect of this sophisticated PI is a faster procurement cycle with greater sharing of knowledge in real time and preservation of knowledge. PI has the potential to help ProdCo realize its business objective of rapid, accurate procurement.

You might imagine, based on robust productivity functionality shown in Figure 11, that even without integration of the underlying enterprise EDA applications, you would be ahead of the game. Having solid connectivity between stakeholders in real time, and streamlined access to documents and knowledge is a big boost to productivity, even without tight integration with ERP and CRM. However, as you start to tie the two sets of systems together, the benefits become striking – assuming of course that the organization and users are up to the challenge of making it all work.

EDA/PI Integration Requirements

What will it take to integrate ProdCo's PI and EDA? To keep it simple and focused, so we can learn, we will look at the functional requirements for EDA/PI integration as they relate to the procurement example only. The following are the requirements that would drive improved productivity in the procurement process through connection between the underlying EDA-enabled ERP and CRM applications and the productivity infrastructure. We will also assign this hypothetical project a name. We are considering integrating ProdCo's PI with its EDA, so the name PIEDA fits well. It sounds a bit like a geek fraternity, which in a sense, it is.

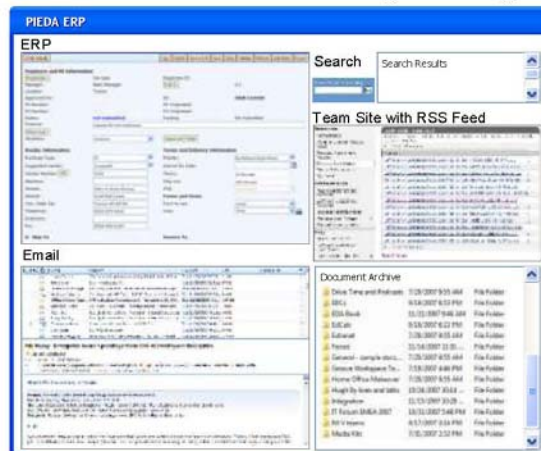


Figure 11

Wireframe mock-up of a composite UI for PIEDA, showing a single interface that contains ERP features, search, email, team site with RSS feed, and a document archive

One of the first requirements that the PIEDA team will have to figure out is the interface. There are two basic choices: integrate the productivity tools into the EDA-enabled ERP and CRM applications, or surface the ERP and CRM applications through the productivity infrastructure tools. Both options require the use of APIs and custom tooling for implementation, though neither requires creating interfaces wholly from scratch. The major productivity suites are available with APIs and development kits that enable integration of interfaces and routing and transformation of messaging to and from enterprise systems.

The first option, which is shown in Figure 11, puts the EDA-enabled ERP app into the same UI as email, document archive, search, and a team site with RSS feeds. In this kind of unified dashboard, the procurement staffer can work on specific RFQs and have a view of his or her email, relevant documents and team updates without leaving the ERP app. In the second option, the ERP app might appear as a side bar in the email client, for example. The best practice for this entire issue might be to do some research first, consulting the end users and showing them the mockups to get input on how they prefer to work before committing to one approach or another.

As shown in Figure 11, one of the most basic features of PIEDA is the ability to automate notification of procurement process status changes to stakeholders. When an order is received from the CRM system, the end user is notified, perhaps through an email alert. Same thing when the "Create BOM" process is executed in the ERP system - the procurement person is notified, and so forth, through Check Inventory, and Create List of Needed Materials. This real time (or rapid) notification of end users of changes in state in the EDA serves to prompt action on the part of the end user. For example, if the procurement staffer is alerted that an order is going into the "Check Inventory" state, then he or she can consult the team site for background information on this type of order and be ready to handle the vendor selection right away. In this way, PIEDA connects the "enterprise nervous system: of EDA right to the human thought processes necessary for completion of the process flow.

Following the process flow shown in Figure 11, the next requirement for PIEDA we need to address is the request for action that occurs when the process reaches the "Request Procurement" stage. When the bill of materials has been generated, and the inventory checked, and the list of needed materials drawn, the procurement is now ready to begin. PIEDA will alert the procurement staffer that he or she needs to create an RFQ and solicit bids. At first, this would probably be an email that requests that action be taken. To make the process efficient, though, the email should contain an embedded link that will take the end user to exactly the right screen in the ERP app where the RFQ can commence. If the end user simply receives an email notifying him that he needs to go to the ERP system and start an RFQ, and making him look it up (perhaps by copying and pasting a job number into a search field) we have not accomplished very much in terms of efficiency.

This innate linking from email, blog entries, IM texts, and documents through to the actual job page on the ERP system is an essential requirement to make PIEDA worth the effort and expense. Users should be able to toggle effortlessly between PI and EDA without having to look up the job detail in the middle of the process.

As the RFQ and bids go through the review loops, PIEDA needs to keep decision makers close to each other and to the related documents. The ERP app needs to show the presence of stakeholders so that all the people involved in the process can connect with one another – either through IM, email, phone, or by looking up information that was authored by a particular person or team. In other words, if a procurement staffer receives an alert that he must review a draft of an RFQ, the draft of the document should appear in the same interface as the alert. Then, the draft should show the presence of its authors, with their availability for IM, email, or phone instantly visible. Alternatively, a user should be able to link on the "mysite," teamsite, or blog of any document author or group of authors in order to learn more about what they know about the procurement process.

It should be possible in PIEDA to conduct a search through the documents, blogs, wikis, teamsites, mysites, and ERP business data to find relevant information about the procurement, or the project itself. Through an enterprise search interface, and back-end enterprise search engine, the end user needs to be able to look up any missing information about the procurement. In our example, if the procurement staffer searches for the component that needs to be shipped in a refrigerated truck, he or she should see the pertinent notes about shipment in the search results –

or at the very least, get a search result that points to documents that contain the needed information.

A more sophisticated version of the enterprise search scenario described above involves including people in the search results. If the end user conducts a search for the bill-of-materials components, the search results could contain both a list of documents related to the components as well as the authors, including their presence and knowledge contributions to the organization. That way, if a procurement staffer cannot find exactly what he is looking for, he can click on the results of a people search and connect to expertise either live or through a published knowledge contribution such as a blog or wiki.

Business intelligence should be a requirement for PIEDA. Though not perhaps a drop-dead necessity, BI gives stakeholders the ability to analyze data and create reports that lead to organizational knowledge. For example, it may not be known that certain component gets destroyed by heat in transit. If a procurement staffer ran a report on job orders that exceeded planned costs, he or she might notice that certain types orders that contained a specific component (which melted in transit) all resulted in poor financial results for their respective orders. Of course, such BI functionality already exists in many ERP systems. The challenge for PIEDA is to make it available to users easily, through the portal front end of the productivity infrastructure. This availability helps expose the knowledge potential of the business intelligence. And, going further, with RSS feeds and subscriptions, it becomes possible to alert people of new knowledge without the receivers of the knowledge needing to know in advance to look for it. Finally, with an EDA-PI connection, the RSS feed itself becomes an event with the potential to trigger action.

The collaborative workflow for document creation needs to be manifest through PIEDA. As end users create documents and revise them, they must be able to see who has contributed to the drafts and be aware of their presence if they need to contact them. The document management functions of PIEDA need to be able to route documents along approval paths, allowing for final approvers to be aware that drafts have been created for their review. The ability to set up an instant online Web meeting or conference call to review a document in real time should be built into PIEDA.

Bottom line, PIEDA needs to provide users with real time (or near real time) access and awareness about events that occur in the ERP and CRM systems. PIEDA needs to give users access to the people who create documents and data inputs to the system. The goal is streamlined, rapid decision making that results in the best possible decisions. PIEDA needs to stimulate the creation and distribution of knowledge in an effortless manner. Users need to find or receive knowledge passively.

PIEDA's Target Architecture

Having outlined PIEDA functional requirements, the challenge now is to relate them to a target architecture for the development of a working EDA. After all, the goal of PIEDA is to reduce the amount of random and unstructured communication between people as they manage the information in the CRM and ERP systems. Doing this means connecting events in those systems to automated, and semi-automation actions in the productivity infrastructure. The first step in this process is to understand how the functional requirements map to events in the overall architecture.

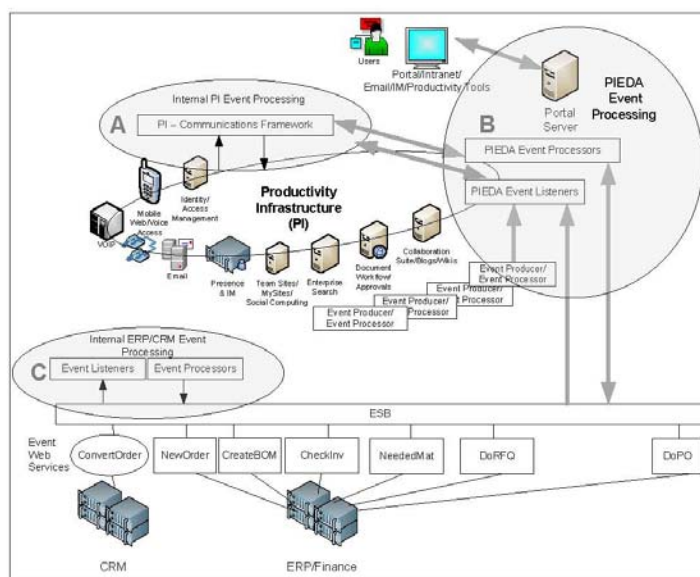


Figure 12

Target architecture for PIEDA, showing three separate areas of event processing: The CRM and ERP systems cycle event data between themselves (area C), as does the productivity infrastructure (area A). PIEDA (Area B) listens for event data from both, and produces its own events. All three event processing areas listen for each other's events.

Figure 12 shows a target architecture for PIEDA. Notable is the contrast to earlier EDA examples we have explored, which have one area of event processing. PIEDA has three. There are events produced, listened for, and processed at the level of the CRM and ERP systems, shown as Area C in the figure. In addition, the productivity infrastructure itself has a whole event processing setup, shown as Area A in the figure. This is an interesting discovery for anyone getting under the hood of a productivity suite: It has many event-driven features right out of the box. Whether you're dealing with Microsoft SharePoint and Office, IBM Lotus, or others, you will find a handsome complement of event processing going on inside the application suite. For example, SharePoint has numerous automated notifications of document draft changes and blog posts, and so forth.

The third area of event processing – Area B – is PIEDA's actual integration between areas A and C. As depicted in Figure 12, PIEDA's event processing area is contained within the portal server. However, it need not be. We are approaching PIEDA as if it were going to be developing using custom development APIs built onto the portal server. When undertaking a project such as PIEDA, many different alternatives may in fact be more attractive. For example, it might make the most sense to build PIEDA on top of the ERP-CRM service bus or develop it on its own standalone application server. The bottom line, though, is it will have to be built. At this time, there is no out of the box solution for the kind of functionality envisioned for PIEDA.

Figure 12 is a rather complicated picture, unfortunately. However, though PIEDA's A-B-C setup is complex, the beauty of it is that areas A and C can operate quite well on their own, even

if you take out B, the PIEDA event processing center. PIEDA is an incremental upgrade to the architecture, and the event processing capabilities of the PI and the back-end are not reliant on it. I mention this because some might look at Figure 12 and decide that the effort is not worth the complexity. On the contrary, the productivity gains should justify the work of making it happen. Plus, if it's done right, the complexity of PIEDA need not result in an excess of administrative load or inhibitions of process agility.

To illustrate how the target architecture for PIEDA fulfills on the functional requirements, I'll walk through one specific example the event flow. Using the requirement of "Request Procurement," let's examine the specific EDA functions that PIEDA needs to realize in order to make the requirement work. The requirement specifies that the ERP system inform that procurement staffer that procurement is needed for a list of items. The following flow of process steps describes how this requirement could be fulfilled, adding in the events that are produced and listened to at each step.

Step	Event Producer(s)	Event Listener(s)	Event Processor(s)	Reaction	Comment
Notify procurement staffer that procurement is needed	DoRFQ (Area A)	PIEDA (Area B)	Email Server (Area A)	Send email that carries event state: Order Number SKUs needed	Email should contain dynamic links to underlying documents and presence information for stakeholders
IF collaboration is needed, create IM session to discuss RFQ	PIEDA Portal Server (Area B) CreateIM	PIEDA (Area B)	IM Server (Area A)	Create IM session that is populated with event state information	Requires XML carried on SIP
IF documents are needed to support procurement, proactively link staffer with documents	PIEDA Portal Server, (Area B) working through Word Processor – PostDocument	PIEDA Portal Server (Area B)	PIEDA Portal Server (Area B)	Publish document update in document library on PIEDA	Document needs to contain author presence data, and links to author mysite and teamsite
Generate vendor solicitation emails	PIEDA Portal Server (Area B) CreateVendSol	PIEDA Portal Server (Area B)	Email Server (Area B)	Create email, populated with RFQ data	RFQ data should be in state carried through event publication

Table 1

Connecting event producers, listeners, processors, and reactions to process steps in "Request Procurement"

Once again, our old EDA friend – carrying state – surfaces in PIEDA. In order for an email or IM session to be generated as a reaction to an event, the event processor must have access to the event state data in the event message. For example, if the user initiates the CreateIM command, which starts an instant message session with another stakeholder, the requirement is that the IM session will automatically contain data about the specific procurement under discussion, as well as links to the ERP system files as well as documents, that need to be discussed. To do this, the CreateIM command needs to contain a function that lets it search for already published event data about the procurement in question. In effect, CreateIM is a two-state command series that looks like this:

- 1) WHEN user initiates CreateIM command for procurement XYZ, SEARCH message queue of events published for DoRFQ and FIND procurement XYZ data
- 2) INITIATE IM Session (through SIP), carrying XYZ data as XML bound to SIP

I use this IM example for a reason, namely that SIP to XML integration is not easy, and the standards involved continue to evolve. Binding SOAP to SIP also challenging, to the point where some sources say it is not possible, or reliable. However, this is exactly the kind of integration that is needed to make EDA-PI integration a success. And, the integration needs to be flexible, allowing low friction for rapid changes in configuring IM session generation with changing business process models. It will take more than average effort to get it to work, but PIEDA should generate greater than average ROI.

The broader point is that PIEDA – area B – needs to be a flexible switchboard connecting the event activities with Areas A and C in order to create ROI and justification of the whole project. If PIEDA is too rigid, and changes are time consuming or expensive, then the EDA PI integration will not serve its business purpose. While this statement could be made about almost any application or architecture project, it is particularly true for PI. PI is inherently more unstructured, more unpredictable than conventional IT situations.

Implementation

Given the complexity of PIEDA, achieving the high degree of flexibility envisioned above is a double challenge. Not only does your development team have to create an architecture that is wholly new, and do so with custom development, they must create an architecture that is highly flexible too. Some wise people might declare the challenge not worth it at the outset.

While these wise people wouldn't be entirely wrong for wanting to avoid the hassle of entering uncharted waters, the good news is that PIEDA can actually be deployed in very small increments. Looking at Figure x, you can understand that Area A already exists in some form. Virtually every sizable business in the world has an email system, an intranet of some kind, and suites of productivity applications. VOIP and corporate IM are on the ascent as well. Building Area A from scratch is not an issue. Area C exists in its component parts, and exposing Web services and creating an EDA is a decision that is separate from building PIEDA. There would be a good rationale for creating an EDA at the CRM and ERP system level for its own sake.

The key takeaway for thinking about implementing PIEDA is to understand that Area B can be developed in stages. In fact, it can be deployed one feature at a time after a core set of EDA infrastructure pieces have been put in place. Unlike Area C, which needs a certain number of event producers, listeners, and processors to hit critical mass and function, PIEDA can start quite small. PIEDA can take one instance of event processing, such as "Request Procurement" and put it into full effect. This approach might even be the optimal way of getting it off the ground.

Given the fickle nature of human-machine interaction, which is truly at the heart of PIEDA, it might be smart to design, test, and deploy in tiny increments. While the dev team might think it's really cool to have an automated IM session generated, users might scoff at such a feature. To save time and resources, and assure the highest level of success, PIEDA's implementation plan should include a thorough usability testing and feedback cycle. End users need to be included intensely throughout PIEDA's lifecycle. This is especially relevant because PIEDA does not rely on pre-packaged software, which usually undergoes its own round of usability tests and market research. In the future, though, PIEDA type platforms might become common. Today, though, it's an unexplored frontier. (Free tip for readers: If anyone is brave enough to stake a claim in this space, venture capitalists might find it interesting..)

Again, in the spirit of not repeating what has already been discussed in great depth previously, let's touch briefly on governance. Governance is where the complexity of PIEDA can be a challenge. With PIEDA, you are talking about integrating EDA components on multiple platforms, with all the federation and mediation that involves. And, you will confront the fact that productivity infrastructure typically comes with its own built-in identity management apparatus that you can't really switch off. It may be possible to adapt the PI's identity management system for use in the broader EDA. Or, you may be able to federate the identity management and governance solutions. Lifecycle and provisioning will also be built in to the PI, and it is almost certain that you will need to keep those functions in place, despite the fact that you will not be able to use those functions for life cycle and provisioning of the EDA components. All of these challenges can be solved, however, with time and effort. In planning for PIEDA, assume that governance will be a major investment.

Conclusion

We are at the very start of the movement towards sophisticated integration of EDA and PI. Early examples of this technological paradigm include the Microsoft-SAP Duet product, which integrates Outlook email with SAP applications, or the Lotus Sametime integration with WebSphere-based business applications. An interesting trend in the evolution of EDA-PI is the use of the "cloud" in linking events and productivity tools. Early examples of this include Facebook.com's ability to create discussion threads and mobile alerts based on postings to a user's page by other network members or Salesforce.com's connectivity with the Outlook email client, taking advantage of Web Services.

It is likely that more sophisticated EDA-PI integrations will continue to develop as the workforce becomes more accustomed to pervasive, mobile computing as a mode of work. The need for disparate groups of people to communicate smoothly on fast-moving business issues regardless of location or computing platform should drive ever increasing capabilities in EDA-PI solutions. This is all the more relevant as a younger generation, who have known nothing other than mobile computing, arrives in the workplace and expects to conduct business in real time across a myriad of temporal and spatial factors.

Mathiesen, Kristy Kay

ORGANIZATION(S):

School of Information Resources and Library Science, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Indigenous Peoples' Rights to Culture and Individual Rights to Access

Overview

This paper uses the methods of philosophical analysis and concepts from moral and political philosophy to defend Indigenous Peoples' human right to control access to their cultural information. As any claim to control access is a threat to the individual right to access, such claims require a justification. The concept of a "right" is delineated and different types of rights are differentiated. The moral foundations of the individual right to access information are articulated. The method of an "overlapping consensus" is used (Rawls 1982). The goal of this method is to find a point of overlap between the various perspectives of persons within a pluralistic society, so that each person can accept a moral principle from within their own "comprehensive conception of the good." In attempting to articulate the moral foundation for Indigenous Peoples' rights to culture, we can start with justifications for limiting access that are already widely accepted within our society. These widely accepted justifications for controlling or limiting access to information include the claim that said information is (a) under copyright, (b) a trade or state secret, (c) harmful to some segment of society, or is (d) private or confidential. The moral grounds for each of these limitations on free access is explored. (a)-(c) are found to not provide an adequate ground for Indigenous Peoples' rights to control access to their culture. The cultural rights of Indigenous Peoples are defended as a form of group privacy. Two possible moral dangers of this defense of indigenous peoples rights to culture are considered: (1) that it would license a sweeping right for all groups to limit outsiders' access to information and (2) that group privacy would allow groups to engage in the abuse of the human rights of members within the "private sphere." It is shown that on a properly nuanced understanding of the contours of the Indigenous Peoples' rights to group privacy, they do not give rise to either of these dangers.

Extended Abstract

On September 13, 2007 the United Nations General Assembly adopted the Declaration on Rights of Indigenous Peoples (UNDRI). The UNDRI asserts group rights of cultural possession and control, which should be of central interest to those in the field of information. According to Article 31 of the UNDRI, for example, Indigenous Peoples have rights to "maintain, control, protect and develop their intellectual property over...cultural heritage, traditional knowledge, and traditional cultural expressions" (UN 2007). And, according to Article 15, Indigenous Peoples have the right to redress for or restitution of "cultural, intellectual, religious and spiritual property taken without their free, prior and informed consent or in violation of their laws, traditions and customs" (UN 2007). These provisions express the ideals behind a world-wide movement by Indigenous Peoples to reclaim their cultural heritage (Brown 2003; Nason 2001).

The question for those interested in the preservation and free flow of information is how to balance the rights of indigenous peoples to culture with the rights of individuals to access information. In order to determine where this balance might be we need a deeper understanding of the grounds of the rights of indigenous peoples over their cultural information. Only if we understand the justification for and the function of these rights can we understand fully what they may reasonably demand of us. Providing such a justification is the goal of this paper.

The first part of the paper lays out some basic theoretical concepts from moral and political philosophy, in particular, it discusses the method of "overlapping consensus" and the concept of a "right." As many have noted, indigenous peoples have unique perspectives on the world that provide an important alternative to western perspectives. Given that there is often a difference of worldviews between indigenous and non-indigenous peoples especially with regard to group vs. individual rights, it is worthwhile to consider how we can go about adjudicating the balance between these rights in a way that can be reasonably justified to all. In "The Idea of an Overlapping Consensus" (Rawls 1982) John Rawls argues that in a pluralistic society like ours, we cannot expect that people will share the same values, metaphysical or spiritual perspectives, and beliefs. He argued, however, that in such a society we can agree on shared principles, via an "overlapping consensus." We can each support a principle from our own perspective, while understanding the perspectives of the others. With this goal in mind, I explore how and whether it is possible to find reasons that we can offer to persons committed to free flow of information that they ought to recognize the right of indigenous peoples to control some aspects of this flow.

Both those who champion individual rights to access and use information and those who assert the rights of indigenous peoples over information use the language of “rights.” Most simply, a right as an entitlement, which others ought to respect (Wenar 2007). These entitlements may be ones with regard to our selves; we may be entitled to do, or not do, something (or be in some state). These entitlements may be ones in regard to others; we may be entitled that others do, or not do, something (or be in some state) (Wenar 2007). Rights can be grounded in the law (legal rights) or in morality (moral rights). A notable type of fundamental moral right is a human right. A human right is a right to that which is necessary for a person to live a minimally good life (Nickel 2007). Using a Hohfeldian (Hohfeld 1919) analysis we can further divide moral rights into claims, mere liberties, immunities, and powers. For our purposes here the most relevant type of right is a claim right. Claim rights impose duties on others. Such duties may either be negative—others have a duty to refrain from doing certain things or they may be positive—others have duty to actively do certain things (Wenar 2007). Both individual rights to access information and group rights to control access to cultural information are moral, negative claim rights. The right to access imposes a duty on others to refrain from interfering with access, while the right to cultural information imposes a duty on others to refrain from accessing certain information. Thus, the greater the individual’s right to access information, the lesser the group’s right to restrict access to information, making clear the need to determine the appropriate balance between these two rights.

Even in a society that claims to favor free speech, citizens accept that there may be some justifications for limiting the free flow of information. We can roughly categorize the rationales for limiting speech as appealing to (a) copyright (or intellectual property more generally), (b) secrecy (including state secrets and trade secrets), (c) harmful material (including such things as obscenity and child pornography), and (d) privacy or confidentiality. In the full paper, (a)-(c) are discussed at length and it is shown that there are serious drawbacks in using any of these rationales as a paradigm for justifying indigenous people’s rights to control access to their cultural information. The right to privacy is found to be more promising.

The right to privacy is often grounded in the value of autonomy (Bloustein 1964). Autonomy is the capacity to “rule oneself”—to be one’s own person and to act from one’s own values and desires without unwanted attention from or interference by others (Christman 2003). But, it is not only individuals, but also groups that may need their autonomy to be protected. We can call this “collective autonomy.” This collective autonomy is violated when others invade the cultural privacy of a group and appropriate cultural information without permission. As Loretta Todd describes it, “Appropriation occurs when someone else speaks for, tells, defines, describes, represents, uses, or recruits the images stories, experiences, dreams of others for their own” (Todd 1990, 24).

If we accept the idea that indigenous groups have a right to cultural privacy, which places a duty on states and individuals to limit access to certain sorts of cultural information, we are faced with two troubling questions. First, would the argument above license a general right of all groups to limit outsiders access to information (Pogge 1997)? Second, would putting all aspects of group life in the “private sphere” of the group make room for abusive power relations within the group (Okin 1999)? In response to the first question, we should note that the special character and context of indigenous peoples—in particular their current situation due to past genocide and oppression—makes their case unique. In response to the second question, it is noted that there are limits to the right of group privacy. Such group rights in no way abrogate the individual rights of the members of the group. As Article 46 of the Declaration of the Rights of Indigenous Peoples states, “The provisions set forth in this Declaration shall be interpreted in accordance with the principles of justice, democracy, respect for human rights, equality, non-discrimination, good governance and good faith.”

References

- Bloustein, E. 1964. “Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser.” *New York University Law Review* 39:962-1007.
- Brown, Michael 2003. *Who Owns Native Culture?* Cambridge, Massachusetts: Harvard University Press.
- Christman, John 2003. 'Autonomy in Moral and Political Philosophy'. In *The Stanford Encyclopedia of Philosophy*. Zalta, Edward N. (ed).
- DeCew, Judith. 2006. "Privacy." *The Stanford Encyclopedia of Philosophy* (Fall 2006 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2006/entries/privacy/>
- Hohfeld, W. 1919. *Fundamental Legal Conceptions*, W. Cook (ed.), New Haven: Yale University Press.
- Nason, James D. 2001. 'Traditional Property and Modern Laws: The Need for Native American Community Intellectual Property Rights Legislation'. *Stanford Law and Policy Review*, 12, pp. 255-66.
- Nickel, James. 2007 "Human Rights." *The Stanford Encyclopedia of Philosophy* (Summer 2007 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2007/entries/rights-human/>.

- Okin, Susan Moller 1999: *Is Multiculturalism Bad for Women?* Princeton: Princeton University Press.
- Pogge, Thomas 1997: 'Group Rights and Ethnicity'. In *Ethnicity and Group Rights*. Shapiro, Ian and Kymlicka, Will (eds) New York City: New York University Press pp. 187-221.
- Rawls, John 1982: 'The Idea of an Overlapping Consensus'. *Oxford Journal of Legal Studies*, 7, pp. 1-25.
- Todd, Loretta. 1990. Notes on appropriation. *Parallelogramme*, 15(1), 30.
- United Nations. General Assembly. 2007. *United Nations Declaration on the Rights of Indigenous Peoples*. New York: United Nations.
- Wenar, Leif. 2007. "Rights." *The Stanford Encyclopedia of Philosophy* (Winter 2007 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/win2007/entries/rights/>.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

Indigenous Peoples' Rights to Culture and Individual Rights to Access

Kay Mathiesen
 School of Information Resources and Library Science
 University of Arizona
 Tucson, AZ 85719
 (520) 621-5219

kmathies@email.arizona.edu

ABSTRACT

Using the methods of philosophical analysis and concepts from moral and political philosophy Indigenous Peoples' human right to control access to their cultural information is defended. The concept of a "right" is delineated and different types of rights are differentiated. The method of an "overlapping consensus" is used. In attempting to articulate the moral foundation for Indigenous Peoples' rights to culture, we can start with justifications for limiting access that are already widely accepted within our society. These widely accepted justifications for controlling or limiting access to information include the claim that said information is (a) under copyright, (b) a trade or state secret, (c) harmful to some segment of society, or is (d) private or confidential. The moral grounds for each of these limitations on free access is explored. (a)-(c) are found to not provide an adequate ground for Indigenous People's rights to control access to their culture. The cultural rights of Indigenous Peoples are defended as a form of group privacy. Two possible moral dangers of this defense of indigenous peoples rights to culture are considered. It is shown that on a properly nuanced understanding of the contours of the Indigenous Peoples' rights to group privacy, they do not give rise to either of these dangers.

Topics

Cultural information systems
 Information policy, ethics, and law

Keywords

Cultural property

Indigenous
 Rights
 Privacy

Intellectual Property
 Access to Information

1. INTRODUCTION

Non-Indigenous Americans belong to a national culture that has benefited from their past and ongoing oppression and destruction of Native American peoples and cultures. As such, we share, along with other colonial and majority populations, a collective responsibility to these peoples. Thus, non-native persons, organizations, and governments have good reasons to support any efforts on the part of native peoples to assert their rights. One of the persisting harms done to indigenous peoples is the attempted destruction of their cultures. This attempted destruction has been two-fold. First is the attempted forced assimilation of indigenous peoples into the culture of the colonizers. The second is the appropriation of that culture through unethical anthropological practices, as well as the

collection and fetishization of native cultures as representations of the exotic, authentic, spiritual, or savage.

In this context, the movements of Indigenous peoples around the world to reassert their connection to their cultural heritage are a positive sign. This assertion has already affected the practice of archaeology, anthropology, and other disciplines that focus on researching indigenous peoples and cultures. Indigenous peoples' continued efforts to get back control of their cultural heritage has and will continue to have important implications for those concerned with collecting, storing, organizing, and providing access to cultural information, e.g., libraries, archives, and museums.

On September 13, 2007 the United Nations General Assembly adopted the Declaration on Rights of Indigenous Peoples (UNDRI) [1]. The UNDRI asserts group rights of cultural possession and control, which should be of central interest to those in the field of information. According to Article 31 of the UNDRI, for example, Indigenous Peoples have rights to "maintain, control, protect and develop their intellectual property over...cultural heritage, traditional knowledge, and traditional cultural expressions." And, according to Article 15, Indigenous Peoples have the right to redress for or restitution of "cultural, intellectual, religious and spiritual property taken without their free, prior and informed consent or in violation of their laws, traditions and customs." These provisions express the ideals behind a world-wide movement by Indigenous Peoples to reclaim their cultural heritage [2], [3].

The question for those interested in the preservation and free flow of information is how to balance the rights of indigenous peoples to culture with the rights of individuals to access information. In order to determine where this balance might be we need a deeper understanding of the grounds of the rights of indigenous peoples over their cultural information. Only if we understand the justification for and the function of these rights can we understand fully what they may reasonably demand of us. Providing such a justification is the goal of this paper.

2. THE PROTOCOLS FOR NATIVE AMERICAN ARCHIVAL MATERIAL AND RESTRICTIONS ON ACCESS

Recently, a number of Native and Non-Native American Librarians and Archivists collaboratively drafted the "Protocols for Native American Archival Material" (PNAAM)[4]. PNAAM describes in detail what these professionals and community members consider the best practices in the handling of Native American archival materials housed in archives, libraries, museums, etc. Materials covered by PNAAM include recordings and transcripts of such works as songs, chants, personal or family information, oral histories, community histories,

"myths," and folklore; Cartographic Materials of such things as sacred sites or areas, village sites, territories, use areas; and archaeological data, ethno-botanical materials, or genealogical data.

2.1 NAGPRA and Information Objects

Discussions of indigenous cultural rights, particularly in the United States, cannot be understood without some familiarity with the Native American Grave Protection and Repatriation Act (NAGPRA). NAGPRA, which was passed in 1990, requires the cataloguing and repatriation of Native American human remains, funerary objects, and objects of "cultural patrimony." According to the National NAGPRA Report so far, "there have been 360 Notices of Intent to Repatriate published...which account for 118,442 funerary objects, 3,585 sacred objects, 296 objects of cultural patrimony, and 768 sacred objects that are also cultural patrimony"[5].

For the purposes of thinking about the status of cultural information the most important component of NAGPRA is the definition of "cultural patrimony" which the statute defines as "an object having ongoing historical, traditional, or cultural importance central to the Native American group or culture itself, rather than property owned by an individual Native American, and which, therefore, cannot be alienated, appropriated, or conveyed by any individual regardless of whether or not the individual is a member of the Indian tribe or Native Hawaiian organization and such object shall have been considered inalienable by such Native American group at the time the object was separated from such group" [6].

Some have argued that the cultural patrimony section of the law covers information objects, such as songs, descriptions or pictures of rituals, etc. However, the language of it being "an object" implies that it does not cover information, which is not typically described as an "object." We can make a distinction between an object—such as a manuscript describing a ritual, recording of a song, picture of a basket, and the information contained in that object. Thus, even if NAGPRA covers the returning of the object, our question is whether it includes the "return" of control over the information contained in the object. Compare the rights I may have in a painting and the rights I have in any reproductions of that painting. If someone steals the painting from me, I have rights to have the painting returned, but I do not necessarily have any rights in reproductions. The right in reproductions may be with someone else, e.g., the artist, or it may be in the public domain. While NAGPRA provides us with a model of how cataloguing and repatriation works, and thus, provides a guide for how such cataloging and repatriation might go for cultural information, NAGPRA itself does not presently provide a legal

requirement for the cataloguing or repatriation of cultural information.

2.2 The Provisions of PNAAM

PNAAM is particularly laudable insofar as it tries to balance the rights and responsibilities of both tribes and non-tribal information stewards, by providing a list of best practices for both communities. It calls for the creation of relationships between the tribes and archives. Thus, anyone dealing with Native American archival materials must read and seriously consider PNAAM.

To give a sense of the suggestions in PNAAM it is worth quoting some of the recommendations at length. With regard to removal or intentional non-preservation of information objects, PNAAM states that, "Some items, such as a photograph of a sacred ceremony, or object, or culturally sensitive documentation of a burial, should not be preserved forever or may need to be restricted or repatriated to the culturally affiliated group" [4]. With regard to restricted access, PNAAM states that archivists and others ought to, "Respect a community's request to restrict access to and use of materials that describe and represent esoteric, ceremonial, or religious knowledge that is significant to the community" [4]. Clearly, although such material may be in the form of original documents in an archive, it may also be included in published works held by academic and public libraries. Thus, the same reasoning that would lead to restricting access to archival materials which represents such knowledge would lead to restricting access to published works that represent such knowledge.

3. INTELLECTUAL FREEDOM AND THE INFORMATION PROFESSIONS

A number of authors in both philosophy and LIS have argued that access to information is key to our capacity to enjoy our democratic and personal freedoms. Alfino and Pierce, for example, argue that, "information... in the context of guided inquiry... supports the development of personal autonomy and personal agency" [7]. Information Professionals, particularly in the U.S. context, see their profession as devoted to intellectual freedom and the public domain. The American Library Association, for instance, proclaims that, "Intellectual Freedom is the right of every individual to both seek and receive information from all points of view without restriction. It provides for free access to all expressions of ideas through which any and all sides of a question, cause or movement may be explored. Intellectual freedom encompasses the freedom to hold, receive and disseminate ideas" [emphasis added] [9]. These values are also embedded in the United States Constitution. The U.S. Supreme Court has held that access to information is a constitutional right. "The right of freedom of speech and press includes not only the right to utter or to print, but the right to distribute, the right to receive, the right to read (*Martin v. Struthers*, 319 U.S. 141, 143) and freedom of inquiry, freedom of thought, and

freedom to teach (see *Wieman v. Updegraff*, 344 U.S. 183, 195)" [8].

This commitment makes any content-based restriction on access to information is suspect. A possible case might make the possible concerns clearer. Suppose, for example, that a student researching folklore wishes to access a recording of a traditional song held by a library or archive. Perhaps he or she wishes to determine whether or how this story might be related to stories in other cultures. Suppose that, after the library or archive consults with the tribe from which the story originated, the tribe wishes to deny access to anyone engaged in this sort of research. Ought the library to restrict access? What would the Library Bill of Rights (LBR), for example, advise a librarian or other information professional to do?

According to Article three of the LBR, "Libraries should challenge censorship in the fulfillment of their responsibility to provide information and enlightenment" [9] It may appear that restricting access to information because a particular group does not want outsiders to know the information amounts to censorship. Even if such limitations are not "censorship" per se, they may still illegitimately infringe our rights to access information. Indeed, the LBR recognizes that there are a wide range of ways in which our rights of expression and access may be abridged and calls for resisting them. According to Article four of the LBR, "Libraries should cooperate with all persons and groups concerned with resisting abridgment of free expression and free access to ideas." [9]. And Article five of LBR seems to pose some serious problems for some of the best practices listed in PNAAM. Furthermore, if groups wish to limit access to works based on the ethnic origin (e.g., whether or not one is a tribal member), age, or gender of the patron, this may conflict with the requirement of the LBR that, "A person's right to use a library should not be denied or abridged because of origin, age, background, or views" [9].

In order to answer our question of how to balance the rights of indigenous peoples to their culture with the rights of individuals to access information we need to have a deeper understanding of the grounds for these rights. Only if we understand the justification for and function of these rights can we understand fully what they demand of us and what they require. Since the focus here is on the rights of indigenous peoples to control access to information and the rights of individuals to access information, it will be worthwhile to familiarize ourselves a bit with the literature on "rights."

4. RIGHTS

Both those who champion individual rights to access and use information and those who assert the rights of indigenous peoples over this information use the language of "rights." This is most clear in the the United Nations

Declaration on the Rights of Indigenous Peoples, and the Library Bill of Rights. However, it is also a powerful theme in PNAAM, which states, for example, that, “Libraries and archives must recognize that Native American communities have primary rights for all culturally sensitive materials that are culturally affiliated with them. These rights apply to issues of collection, preservation, access, and use of or restrictions to these materials” [4] [emphasis added]. It will be useful, then, to briefly discuss what rights are, and more specifically how we should understand the rights claimed by those advocating access and those advocating the rights to control access and use of information.

Typically when someone says that they have a right to *x*, they are referring to what the rights theorist W.H. Hohfeld called “claim rights” [12]. If I have a claim right to read, for example, that means that others would have correlative duties that they ought to fulfill. If I have a claim right to read a book, then you may have some sort of duty to make the world such that I actually can effectively read. If you fail to do your duty, then you are “violating” my right. Your duty might be to not interfere. So, for example, in asserting my right to read, I may be saying that you have a duty to you refrain from criticizing me publicly for reading the book, turning off the lights so I can’t read the book, or to remove the book from the public library. Or your duty may be to actual in some way assist me. So, for example, others may be obligated to fund public schools where I would be taught to read and public libraries that would stock this and other books. Rights theorists typically call the right imposes for first sort of duty on others a negative (or liberty) right and the right that imposes for second sort of duty on others a positive (or welfare) right [13]. In other words, if I have a liberty right to read, then others are obligated to refrain from interfering with my reading. If I have a welfare right to read others are obligated to provide me the conditions necessary for reading.

Below I will often talk about the “contours” of a right. It is worth saying a bit here about what I mean by that. We often think of rights as simple—you have a right, or you don’t. So, for example, if I have a property right to my car, then we might think of it as a sort of bubble around my car—I can do what I want with my car—dent it with a hammer, paint it bright pink, and others cannot do things to my car without my permission. But rights are almost never so simple. There are all sorts of things that I am not allowed to do to or with my car—I cannot drive it 100 miles an hour on a city street, I cannot drive a car that pollutes, etc. Thus, my rights to my car do not form a sort of smooth circle around my car, but rather has contours of what I can and cannot do in relation to my car.

Consider my “right to read” again. We often say that we have a constitutional right to read what we wish—again it

is as if there is a bubble of protection around our acts of reading that forbid others from interfering with them. But, this bubble is dented with all sorts of exceptions. You do not have a constitutional right to read other’s private medical records, to read pirated copies of a work, to read documents that contain national secrets, or to read slanderous remarks about someone. In all these cases the government is permitted to pass laws to restrict our access to such works. Our rights to read have contours that are quite complex. Given this, the mere fact that one has a “right” with regard to *x*, does not answer the question as to whether one can do some particular thing with regard to *x*. It will depend on the precise contours of the right in question. In particular, exactly what actions does the right authorize?

With this account of rights at our disposal we can characterize what the potential conflict between claims of rights to culture and claims to freedom of access. On the one hand, there is the view that indigenous peoples have a right to control access to cultural information. This means, that, if they wish, they may assert a right that, with regard to certain information, it be destroyed, all copies be returned to the group, or that, while the information objects may remain outside the direct physical control of the group, access this information should be restricted according to the wishes of the group. Such a right would place a duty on others to refrain from accessing this information and may place a duty on the current “stewards” of the information to ensure that the rights of indigenous people with regard to this information are respected. Since the rights of indigenous peoples place a duty on others not to access information in some cases, this means that outsiders would not be free to access such materials.

5. OVERLAPPING CONSENSUS

Many have noted that indigenous peoples have unique perspectives on the world that provide an important alternative to western perspectives. As Ermine notes, western and indigenous peoples have, “disparate worldviews each formed and guided by distinct histories, knowledge traditions, values, interests, and social, economic, and political realities” [14]. Indeed, this difference is often cited as an important reason that such cultures ought to be protected and preserved. However, the fact that indigenous peoples have unique world-views creates a special challenge to any theorist trying to provide a discussion of the reasons that might support their rights claims. Indigenous peoples themselves may have unique beliefs about the world and their place in it that grounds their claims to various cultural rights. Non-indigenous peoples should recognize the importance that such beliefs have for indigenous peoples and, ideally, could learn new and possibly better ways of being in the world. Nevertheless, if as a matter of fact non-indigenous peoples do not share these views, then the reasons that may be

compelling for indigenous peoples will have no grip on non-indigenous peoples.

John Rawls was the preeminent Political Philosopher of the late 20th and early 21st Century. Rawls developed a theory of how principles of justice might be grounded and justified without appealing to what he called “controversial metaphysical commitments.” Rawls argued that in a pluralistic society such as many of us live in, we cannot expect that people share the same values, metaphysical or spiritual perspectives, and beliefs [15]. He argued, however, that in such a society we can agree on shared principles, via an “overlapping consensus.” We can each support a principle from our own perspective, while understanding the perspectives of the others. A simple example from the United States context will illustrate this. Consider the idea of the separation of church and state. Both religious and non-religious people can accept this principle as expressing the idea of the freedom of conscience. However, the religious person may be committed to such a principle on the grounds that persons should come to God as a matter of individual conscience, which would be interfered with if the state in any way mandated a particular religion or religious instruction. The atheist may be committed to such a principle on the grounds that religions are all equally false, but given that people are personally committed to them they should be free to indulge in their religious beliefs and practices as much as any other private club or activity. The religious person and the atheist do not have to share robust world-views including spiritual, metaphysical, or moral beliefs in order to share a commitment to the principle of the separation of church and state.

It is worth noting that this view is not based on skepticism about whether there is moral or religious truth. Rather, it is based on the philosophical and practical recognition that reasonable persons differ on such issues. The fact that people differ is not based on the ignorance or perfidy of some. Thoughtful and intelligent persons of good will can have very different views about the nature of the universe and the place of human beings in this universe. In the rest of this paper, I try to find principles which can form the basis of an overlapping consensus of both indigenous and non-indigenous peoples.

6. JUSTIFICATIONS FOR LIMITING ACCESS TO INFORMATION

As noted above rights are always complex. To say we have a right to free speech is not to say that we can say whatever we want or publish whatever we want. There are many limitations on free speech with which we are quite comfortable and familiar. We often do not think of them as limitations, because we simply don't think of certain activities as being covered by a right to free speech. Here I am going to focus on the limitations standardly found in U.S. law. We can roughly categorize such rationales for

limiting speech as the following: Copyright (or intellectual property more generally), Secrecy (including Government Secrecy and Trade Secrets), Harmful Material (including such things as Obscenity and Child Pornography), and Privacy.

6.1 Intellectual Property

The most commonly given account of the basis of the putative right of indigenous peoples to control access to cultural information is intellectual property. The rights of authors under copyright include the exclusive right to publish a work, to publicly display the work, to make copies of the work, and to create derivative works based on the original work [16]. If indigenous peoples hold copyright in their cultural information, then it would be prohibited for outsiders to reproduce cultural information, to display it, or to copy it. So, for example, it would prohibit a person from placing a traditional song, image, or story on the internet without the prior permission of the tribe to whom this song, image, or story belongs.

The foundational idea of intellectual property is that authors have rights in their creations. And, in many cases the “author” of a work is not a single individual, but a group of individuals. In such cases the copyright is often held by a collective entity of which these individuals are members. Think, for example, of the copyright that is held by the Disney Corporation. The works in question are often the creation of a large number of persons working together under the aegis of Disney. Similarly, traditional songs, stories, visual works, etc. are the creations of individual indigenous persons working together as members of a particular tribe. Following the logic of intellectual property, indigenous peoples would have authorship rights in their culture, because it is their collective creation.

There are a number of ways in which copyright seems exactly suited to the sorts of rights that indigenous peoples claim in their cultural information. However, there are a number of features of copyright, at least as it has been established in the U.S. constitution and law, that make it a poor fit as a paradigm for the rights that indigenous peoples have in relation to their cultural information.

First, the tradition of copyright in the anglo-american context has always been that it is time limited. Indeed, originally copyright could only be held for 14 years with one possible extension [17]. Even now, while the period of copyright has been significantly extended, to life of the author plus 70 years, or for corporations 90 years after the publication of the work. The idea behind these limitations is that the rights of authors to their cultural property must be balanced against the importance of the free flow of information characterized by the public domain. Even if we argued that all cultural information should now be repatriated based on the argument that past publication of

the works were forced rather than a genuine choice of the peoples in question, any future publication, public performance, etc. of the works would start the clock ticking. Indigenous peoples would likely not be happy with the thought that if they in any way publicly share their culture they will lose rights over it in one hundred years. They precisely do not conceive of their culture as something that can go into the public domain of the larger culture to be used as persons wish.

Second, in order for a work to qualify for copyright, it must be fixed in a “tangible form of expression” [16]. Thus, for cultures that have traditionally had strong oral traditions where works may not be written down, copyright would not cover such cultural information. Indeed, as long as a work is oral and not written, the copyright would belong to the person or group who records or writes down a story or song, not to the person or group who originated the story or song.

Third, the rights of authors under copyright only cover the form of the expression, not the idea or story expressed. With regard to a story, for example, it covers the actual words used to tell the story, not the form or type of story itself. This is even more true for descriptions of practices. While an indigenous group may have copyright on a particular description of a practice that they engage in, they do not have copyright over the description of this practice. And, unless they write down or record the particular movements of the practice, they have no copyright over it if others come to know about it.

Finally, even if indigenous peoples were willing to accept time limits (or even if we were willing to remove the time limits in cases of indigenous cultural information), there are fair use exceptions to all of the exclusive rights that authors have in their works. No author has complete control over his or her works. Others may use or quote from the work for the purposes of commentary and criticism. Others may use small pieces not central to the work in “transformative” uses. And, if the use is for scholarly, as opposed to monetary gain, there is more room for fair use. It is not clear that indigenous peoples would find such exceptions acceptable. It is a mistake to think that indigenous peoples only object to their works being used by others for monetary gain. It is the inappropriate access and use of such information that is the problem and the inappropriateness can equally be in someone who is using the works for scholarly or educational purposes.

6.2 Secrecy—Trade and State

The law also protects the right of certain groups to keep information secret. Some of this falls under the rubric of intellectual property in the form of trade secrets. There are also laws protecting state secrets. Some have argued that trade or state secrets are the right model for justifying and

protecting indigenous peoples rights to their cultural property.

According to the Uniform Trade Secrets Act [18], a “trade secret” is: “information, including a formula, pattern, compilation, program device, method, technique, or process, that: (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.”

There are features of trade secrecy that again seem to make it a poor fit for indigenous people’s rights to cultural information. It is, like other things covered by intellectual property law, intended to promote discovery and innovation. The idea is that people will not be able to spend the time as a group developing new products, inventions, or techniques, if their work can be taken and used by others. In particular, the focus is on economic or monetary value that a secret may potentially have had. Thus, the value that indigenous persons may place on some information being secret is not captured in the current law on trade secrets.

Some have suggested that indigenous cultural information is akin to state secrets. In one way state secrets seem a better fit. They do not focus on monetary value or in promoting innovation. However, they do focus on protecting security. According to the United States Court of Appeals, “The state secrets privilege is a common law evidentiary rule that allows the government to withhold information from discovery when disclosure would be inimical to national security” [19]. It is not clear that all (or even most) the cultural information that indigenous peoples would want to protect would fall under the concept of a state secret necessary to protect security. If it is protect the security of their land, then this may give indigenous persons an internal reason to have customs, policies, or laws protecting such information. This would limit the information so protected to that which could be legitimately understood as necessary to protect the security of the indigenous people in question. Furthermore, the idea that certain information is “secret” depends on the fact that they have not been revealed. There seems to be little sense in saying that something that has been published and has been widely available in libraries and the internet is “secret” information. Thus, an appeal to secrecy would not provide grounds for repatriating information.

6.3 Harmful Material

In addition to intellectual property and secrecy limitations on access to information U.S. law also has provisions to

limit people's access to "harmful" information. Information may be harmful in the sense that a person can use the information to engage in harmful activities. So, for example, we might limit people's access to information about the security plans of a power plant or town reservoir. Information may be considered harmful in the sense that it is harmful for people to be exposed to certain information. So, for example, there are limitations on how much graphic violence can be shown on broadcast television. The idea is that it can be upsetting to people to be unwittingly exposed to such material. And, some may even think that information may be harmful to some (children), but not necessarily harmful to others (adults). Thus, the law limits children's access to certain sorts of explicit sexual material, while not limiting adults' access to such material.

It may be argued that certain indigenous cultural information can be harmful if accessed by the wrong person in the wrong way. While arguments that information may be harmful are frequently a subject of quite strong disagreement and debate, decisions about whether material may be restricted due to putative harm is standardly determined through democratic processes as constrained by the Constitution. It is very possible that non-indigenous persons will not share the views of indigenous persons about the harm that could result from exposure to songs, images, or information about religious practices. In other words, the harm justification is unlikely to gain an overlapping consensus.

6.4 Privacy and Confidentiality

The final justification for limiting access to information is based on claims of privacy or confidentiality. Informational privacy concerns information about persons, which is considered "sensitive." According to Ronald Standler, for example, "Privacy is the expectation that confidential personal information disclosed in a private place will not be disclosed to third parties, when that disclosure would cause either embarrassment or emotional distress to a person of reasonable sensitivities. Information is interpreted broadly to include facts, images (e.g., photographs, videotapes), and disparaging opinions" [20].

On the face of it, privacy may seem to be the least likely candidate for grounding indigenous people's rights to cultural information. One tends to think of privacy as essentially an individual right. Indeed, a number of authors have specifically characterized the right of privacy as "the right to be let alone" [21]. And the supreme court, in rejecting the claim of a corporate right to privacy held that "corporations can claim no equality with individuals in the enjoyment of a right to privacy" [22]. Indeed, some have actually characterized a right to privacy as a right to "individuality." This seems on the face of it to exclude

groups such as tribes or clans from defending their claim to their cultural information as a type of right to privacy.

However, I want to argue that indeed there is a notion of group privacy and that it can provide the best grounds for indigenous peoples rights to limit access to their cultural information. The value of the right to group privacy—to control what information those outside of the group have about the group—is essentially the same value as individual privacy. We can find three accounts of the underlying values that a privacy right serves in the literature on privacy. We can call these strands the Consequentialist, the Kantian Liberal, and the Communitarian views. All of these views provide support for a group right to privacy in cultural information.

6.4.1 Consequentialist Defense of Group Privacy

On the consequentialist view the value of privacy is that it allows us to gain certain benefits and avoid certain harms. Posner discusses this feature of privacy in an article that in general is skeptical about the value of privacy [23]. He argues that certain sorts of goods—e.g., those that require confidentiality, could not be enjoyed without privacy. Some argue that the fact that people find invasions of privacy personally distressing provides at least some moral grounds for preventing and avoiding such invasions.

It is obvious that the group's ability to control access to their cultural information will often have good consequences. First, in line with Posner's argument, it will make them more willing to share information among themselves and with outsiders, if they know that they can control the extent to which such information is disseminated. This is valuable for the ability of the group to have a rich culture and allows a freer flow of information to the outside. If the group has no ability to control the dissemination of information about their culture, they may be less likely to share such information for fear that it will leak out into the public domain. Second, it is clear that many indigenous peoples find it distressing for outsiders to know about, publicly display, etc., cultural information. Thus, limiting other's ability to do this will avoid the bad consequences of this distress.

6.4.2 Kantian Liberal Defense of Group Privacy

The Kantian Liberal view holds that privacy is essential to human dignity. In particular, it is argued that privacy provides a sphere of autonomy for individuals which is necessary for them to live authentically human lives. If other's can use personal information about me for their own ends, they fail to treat me with dignity as a human being [24]. And, if other's can observe me without my permission, they interfere with my ability to make my own choices free from external scrutiny. I call this view the Kantian Liberal view of privacy, because it's emphasis on the importance of individual autonomy is clearly inspired by Kant's idea that persons are fundamentally rational

beings capable of choice and his categorical imperative that persons ought not be treated as a means. It is liberal insofar as it focuses on the liberty of the individual and the fundamental grounds of our capacity for individuality and individual expression.

It is not only individuals, but also groups that may be “autonomous.” We can call this “collective autonomy.” Collective autonomy is closely tied to the concepts of self-determination and sovereignty. In order to have collective autonomy, we may argue that a group needs what Loretta Todd has called “cultural autonomy.” According to Todd, cultural autonomy “signifies a right to one’s origins and histories as told from within the culture and not as mediated from without” [25]. This right to cultural autonomy is violated when others appropriate cultural information. As Todd describes it, “Appropriation occurs when someone else speaks for, tells, defines, describes, represents, uses, or recruits the images stories, experiences, dreams of others for their own” [25]. This idea is closely tied to the argument that Bloustein [24] makes that personal dignity is harmed when persons are used as a “means” to others entertainment, curiosity, or economic gain. In such a case, a person’s dignity and ability to control his own life—by having control over his identity, are harmed. In a similar way, a group that cannot have some control over how its identity as defined when its culture is used, exploited is treated as a means, rather than as an end in itself.

6.4.3 Communitarian Defense of Group Privacy

The Communitarian view of the value of privacy holds that privacy is essential to us as social beings [26] [27]. In particular, it allows us to modulate how much and what type of information that others have about us. This allows us to form intimate relationships by selectively revealing private information to some persons. According to Charles Fried, for example, intimacy requires that we have the ability to reveal information about ourselves, which we do not reveal to all. Privacy is thus a means to modulate degrees of friendship ([26], 210-211) Privacy provides freedom to define ourselves in relation to others, by not revealing our every thought. ([26], 212).

The Communitarian argument can also be marshaled to defend a group right to privacy. Group privacy is important for both the capacity for members within the group to form relationships and engage in various social and cultural activities, and to modulate their relationships with external individuals and groups. The connection between privacy and the ability to form relationships can also be seen in the case of group privacy. Group privacy is necessary for the creation and maintenance of both internal relationships among the members of a group and external relationships between the group and other groups or individual non-group members. Indeed, Gerstein specifically makes this connection in his discussion of

privacy [27]. He notes that from within the experience of a religious ceremony its significance depends on its religious meaning for the participants. “This means that there will always be the threat of the degeneration of the religious act into a mere form, a hollow shell with nothing but its surface attractiveness to give it value. Again, the need to deal with this threat from within makes it necessary to resist observation from without. The presence of the onlooker, if it is not met with resistance, makes those involved in the ceremony accomplices in a concentration on the surface of things at expense of inner meaning” ([27], 270). He argues that even if one’s actions are motivated by noble goals they can be destructive to the meaning of the act for the participant ([27], 270). “[H]e may well feel as sense of desecration whenever what was to function purely as part of the expression of a relationship between himself and his God is made use of by others—even as a means of learning about that relationship” ([27] 270)

7. LIMITATIONS ON GROUP PRIVACY

If we accept the idea that indigenous groups have a right to cultural privacy, which places a duty on states and individuals to limit access to certain sorts of cultural information. we are still faced with an important question. “Why do Indigenous Peoples (in particular) have a right to cultural privacy?” What about other groups such as Scientology, the Catholic Church, the American Library Association, the Amish? Do they also have rights to cultural privacy that we have a duty to respect?

There are a number of features of the case of indigenous peoples that make their claim to cultural privacy much stronger. First, recall that the function of privacy protections is to protect and enhance autonomy, individual or collective, and to preserve individuality, personal or cultural. Some sorts of entities have a greater claim to autonomy promoting protections than others—adults more than children, humans more than animals. Interestingly, it is because some entities have a greater capacity for autonomy and a greater need for it that we are more concerned to promote it. Some groups have a greater capacity and need for autonomy than others. The more “comprehensive” a group is—the more aspects of life it encompasses, the greater its capacity and need for autonomy or “self-rule.” Indigenous groups often encompass multiple aspects of their member’s lives—spiritual, physical, educational, political, etc. Other groups, such as the American Library Association, for instance, do not have this sort of comprehensive character. The ALA only encompasses the professional lives of its members. Typically members will belong to many other sorts of groups, which will encompass other parts of their lives—family, religion, political party, ethnic group. Given this the ALA has both less of a capacity to be fully an autonomous group and it has less of an important role in the lives of its members. Thus, such comprehensive

communities as indigenous peoples will require greater autonomy, thus will have a stronger claim to informational privacy.

Secondly, it may be that there are certain contextual features of some groups that provide grounds for greater informational privacy protections. Two dimensions of the context of indigenous peoples give them greater claim to cultural privacy than many other groups: their current context and the historical context of genocide, forced assimilation, and cultural appropriation. In the current national and international context many indigenous peoples are parts of larger states or cultural groups that are both culturally different and culturally dominant, there is a greater need for that group's culture to be protected [28]. The case of indigenous peoples, indeed, is of particular note, because they face not just neglect, discrimination, and swamping by the majority culture, but the active appropriation and exploitation by the majority culture. Many within majority cultures do not just neglect the cultures of indigenous peoples, but try to attain some sort of false authenticity through acts of appropriation. Such acts of appropriation are often also commercializing and imply that the cultures are "dead" or something of the past, rather than living cultures within which people find meaning and live their lives. This appropriation threatens the meanings of the culture for those who live within it. This is a sort of threat that is not faced by such groups as those listed above.

Given that the last two reasons for supporting a duty to respect the right to cultural privacy on the part of indigenous peoples are heavily contextual, it may be that such rights and the concomitant duties, are remedial, rather than eternal. While the concept of cultural privacy is important, it may be that the values of free flow of information are sufficiently important that when a group is culturally strong we should strike a different balance from the one we strike when the group had not had the power to use normal means to protect its culture in the past.

There are, however, important limits to the rights of cultural privacy. It is important to note that indigenous groups do not exist in a realm separate from other groups and individuals, thus it is neither possible nor feasible that they have "absolute" privacy. What they do affects others and insofar as they engage with others, there needs to be some degree of transparency in their interactions and they need to be open to discussion and criticism.

In particular, it is important that we address concerns that privacy claims may be used to cover up corruption, abuses of power, deception, etc. As Judith DeCew has pointed out, "[P]rivacy appears to be something we value to provide a sphere within which we can be free from interference by others, and yet it also appears to function negatively, as the cloak under which one can hide

domination, degradation, or physical harm to women and others" [29]. Privacy claims may be used to block any public criticism by arguing that the criticism includes "private" information about the group. Given this, it will be important to carve out room for a kind of Whistle Blower Exception to the rights of groups to protect their cultural privacy.

Finally, the right of group privacy should not be understood as absolute. There are other rights of individuals as human beings—as encoded in the U.N. declaration of human rights, with which these group rights must be balanced. Thus, if some practice is abusive of individual persons, the right that the group has to privacy is overridden. Of course, some may object to any restriction on the grounds that any outside interference will be tantamount to imperialistically imposing values. However, the presumption of this paper is that there are some shared values to which we may agree. And, insofar, as the indigenous peoples' points of view are represented in the UNDIR, they are committed to the provisions of the U.N. human rights document. It is the view expressed in UNDIR that there is no contradiction between the rights of indigenous peoples and the individual rights as expressed in the UNDHR, which include rights of expression and access to information.

9. REFERENCES

- [1] United Nations. General Assembly. 2007. United Nations Declaration on the Rights of Indigenous Peoples. New York: United Nations.
- [2] Brown, Michael 2003. Who Owns Native Culture? Cambridge, Massachusetts: Harvard University Press.
- [3] Nason, James D. 2001. Traditional Property and Modern Laws: The Need for Native American Community Intellectual Property Rights Legislation. *Stanford Law and Policy Review*. 12, 255-266.
- [4] First Archivists Circle. 2007. Protocols for Native American Archival Materials. DOI= <http://www2.nau.edu/libnap-p/protocols.html>
- [5] National Park Service. U.S. Department of the Interior. 2007. National NAGPRA FY07 Midyear Report. DOI= <http://www.nps.gov/history/nagpra/>
- [6] Public Law 101-601--Nov. 16, 1990 104 Stat. 3049.
- [7] Alfino, Mark and Pierce, Linda. 1997. Information ethics for librarians. Jefferson, North Carolina: McFarland.
- [8] *Griswold v. Connecticut*. 381 U.S. 479. (1965).
- [9] American Library Association. 2006. Intellectual Freedom Manual. American Library Association, Chicago, IL.
- [10] Heins, Marjorie. 2004. On Protecting Children--From Censorship: A Reply to Amitai Etzioni. *Chicago-Kent Law Review* 79.
- [11] Boaz, Martha. 2003. Censorship. In *Encyclopedia of Library and Information Science*, ed. M. J. M. Bates,

- Mary Niles, Drake, Miriam. New York: Marcel Dekker.
- [12] Hohfeld, W. 1919. Fundamental Legal Conceptions, W. Cook (ed.), New Haven: Yale University Press.
 - [13] Wenar, Leif. 2007. Rights. The Stanford Encyclopedia of Philosophy (Winter 2007 Edition), Edward N. Zalta (ed.), DOI=
<http://plato.stanford.edu/archives/win2007/entries/rights/>.
 - [14] Ermine, Willie. 2007. Ethical Space: Transforming Relations. DOI=
http://www.traditions.gc.ca/docs/docs_disc_ermine_e.cfm#ermine
 - [15] Rawls, John 1982. The Idea of an Overlapping Consensus. Oxford Journal of Legal Studies. 7, 1-25.
 - [16] U.S. Copyright Act, 17 U.S.C. §§ 101 – 810.
 - [17] Samuels, Edward. 2000. The illustrated story of Copyright. Thomas Dunne Books/St. Martin's Press, New York, NY.
 - [18] National Conference of Commissioners on Uniform State Laws. Uniform Trade Secrets Act. DOI=
<http://nsi.org/Library/Espionage/usta.htm>
 - [19] Zuckerbraun v. General Dynamics Corp., 935 F.2d 544, 546 (2d Cir. 1991).
 - [20] Standler, Ronald. 1997. Privacy Law in the United States. DOI= <http://www.rbs2.com/privacy.htm>
 - [21] Warren, S.D. and Brandeis, L.D. 1890. The right to privacy. Harvard Law Review. 4, 5, 193—220.
 - [22] U.S. v. Morton Salt Co., 338 U.S. 632, 651-52 (1949) [citations omitted]
 - [23] Posner, R.A. 1978. Privacy, secrecy, and reputation. The University of Chicago. Chicago, IL.
 - [24] Bloustein, E. 1964. Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser. New York University Law Review. 39, 962-1007.
 - [25] Todd, Loretta. 1990. Notes on appropriation. Parallelogramme, 15(1), 30.
 - [26] Fried, Charles. 1984. Privacy: A Moral Analysis. Philosophical Dimensions of Privacy: An Anthology. Schoeman, F.D. Editor. Cambridge University Press, Cambridge.
 - [27] Gerstein, Robert S. 1984. Intimacy and Privacy. Philosophical Dimensions of Privacy: An Anthology. Schoeman, F.D. Editor. Cambridge University Press, Cambridge.
 - [28] Margalit, Avishai; Halbertal, Moshe 2004: 'Liberalism and the Right to Culture'. Social Research, 71, pp. 529-48.
 - [29] DeCew, Judith. 2006. "Privacy." The Stanford Encyclopedia of Philosophy (Fall 2006 Edition), Edward N. Zalta (ed.), DOI=
<http://plato.stanford.edu/archives/fall2006/entries/privacy>

FULL PAPER SUBMISSION 16:

Privacy and Participation in Ubiquitous Information Systems: Information Ethics when Mobile Phones are Sensors

AUTHOR(S):

Shilton, Katie; Burke, Jeff; Estrin, Deborah; Hansen, Mark; Srivastava, Mani

ORGANIZATION(S):

University of California, Los Angeles, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Privacy and Participation in Ubiquitous Information Systems: Information Ethics when Mobile Phones are Sensors

Ubiquitous information systems hold increasing promise for widespread participation in data collection and dissemination. Common and abundant devices such as mobile phones can sense and record data such as location, sound, and images. These systems can facilitate community participation in basic and applied research, improvement of quality of life, social change, self-reflection, and creative expression. But the design and use of these tools also pose new challenges for privacy, data security, and ethics.

This paper explores responsible approaches to the design of systems for ubiquitous digital capture. The authors include an Information Studies researcher working alongside computer scientists and engineers from the Center for Embedded Networked Sensing (CENS). This ongoing partnership pairs ethical frameworks developed within Information Studies and Science and Technology Studies with system design expertise and concrete, real-world challenges. For example, CENS is developing a platform to share activity inferences generated with geospatial data from mobile phones to assess personal environmental impact. CENS is also exploring systems for capturing and sharing data about neighborhood walkability and other community assets. The ongoing research detailed in this paper investigates how such ubiquitous sensing projects can conscientiously collect personal data for social science and participatory research applications.



Figure 1: A CENS project tracks and shares users' movements.

Background

Previous research has proposed a variety of technical approaches to privacy protection in a world of increasing information capture. These include encryption, privacy-enhancing technologies (PETs), and statistical anonymization of data (Burkert, 1998). While these approaches to privacy protection are important in human-embedded sensing systems, we argue that privacy requirements for networks of sensitive personal data extend beyond such measures.

Individuals experience privacy as a fluid concept. Privacy concerns extend beyond anonymity or data protection; privacy can be a process of regulating boundaries or a method of portraying particular personal identities (Palen & Dourish, 2003). Privacy expectations are shaped by social conventions (Waldo, Lin, & Millett, 2007) and by places and spaces of data capture (Cohen, 2008). Individuals may feel threats to privacy during data collection, during data sharing, or if data remains accessible indefinitely.

Designing for Privacy and Participation

This paper details design considerations and practices that promote a balance between participation and privacy. An interdisciplinary team hosted by CENS has iteratively developed design principles that respect a multi-faceted definition of personal and social privacy. The paper will detail design guidelines such as targeted capture, data retention policy, and methods

for identification of privacy pressure points in a ubiquitous information system. It will emphasise techniques for responsive system design based upon user-specified privacy needs. Finally, we will emphasize participation over restriction as a response to privacy ethics. We argue that by finding a balance between privacy and participation, embedded network sensing systems can reach their research, empowerment and documentary potential while respecting participant privacy concerns.



Figure 2: CENS interns participate in cooperative data gathering and analysis.

References

- Burkert, H. (1998). Privacy-enhancing technologies: Typology, critique, vision. In P. E. Agre & M. Rotenberg (Eds.), *Technology and privacy: The new landscape* (pp. 125-142). Cambridge, MA and London: The MIT Press.
- Cohen, J. E. (2008). Privacy, Visibility, Transparency, and Exposure. *University of Chicago Law Review*, 75(1).
- Friedman, B., Kahn Jr., P. H., Hagman, J., & Severson, R. L. (2006). The watcher and the watched: Social judgements about privacy in a public place. *Human-Computer Interaction*, 21, 235-272.
- Palen, L., & Dourish, P. (2003, April 5-10). *Unpacking "privacy" for a networked world*. In Proceedings of CHI 2003, Ft. Lauderdale, FL: ACM, 129-136.
- Waldo, J., Lin, H. S., & Millett, L. I. (2007). *Engaging privacy and information technology in a digital age*. Washington, D.C.: The National Academies Press.

FULL PAPER SUBMISSION 17:
Ethics and the Undergraduate

AUTHOR(S):

Conway, Paul
Carbo, Toni
Kvasny, Lynette

ORGANIZATION(S):

University of Michigan, United States of America;
University of Pittsburgh, United States of America;
Pennsylvania State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Session: Ethics and the Undergraduate

Session Description

Research on the ethical challenges at the nexus of information, technology, and people is a burgeoning international, multidisciplinary effort with an emerging theoretical foundation and a rich scholarly literature. For example, in the area of theory development, Floridi (2007) is constructing a unified general approach to macro information ethics that holds the promise of bridging multicultural perspectives and religious beliefs. His ideas are being adapted and applied to issues as diverse as the digital divide, cheating on games, and environmental stewardship. Focusing on new and emerging technologies, Moor (2005) hypothesizes an intimate association between technological change and ethics. "As technological revolutions increase their social impact, ethical problems increase." His perspective is compelling not just because new technologies are increasing their reach but also because revolutionary technologies create new opportunities for action before well thought out ethical policies have been developed. "We need to learn about the technology as it is developing and to project and assess possible consequences of its various applications." Highlighting the transformative nature of information technologies, Capurro (2006) proposes that the basic moral principle our digital age is "the right to communicate in a digital environment which includes the right to preserve what we communicate for future generations."

Nearly every member of the iSchool Caucus embeds research on information ethics into the curriculum at the graduate level, where the tensions among professionalism, culturally based behavioral norms, and the capabilities of new technologies is most apparent. At the undergraduate level, however, the story may be quite different. Today's undergraduates, the so-called "NetGeneration" whose formative experience encompasses new and emerging technologies, may be through their uses of these technologies reconfiguring the ethical landscape and re-writing the rules of social engagement in a networked world...or not. In teaching ethics and information technology, faculty face the particular challenge of accounting for new research and ethical theories, while bridging generational experience, exploring interdisciplinary exploration of complex ideas from interdisciplinary perspectives, and embracing pedagogical innovation.

This session explores the foundations of the research that underlies variation in the teaching of information ethics at the undergraduate level within the iSchool context. The moderator will establish the context of the papers by reporting on a brief survey of the iSchools Caucus on their undergraduate ethics curricula. Each of the three papers presents a distinctive model for structuring intellectual content and engaging students in an exploration of ethical issues. The models vary in terms of their philosophical orientation, on the balance in emphasis on policy, law, moral principles, and technological determinants of human behavior, and in pedagogical approach. One paper emphasizes innovative uses of social computing tools to expose and examine ethical dilemmas. Another paper juxtaposes information ethics and public policy. The third paper considers social justice as an appropriate conceptual framework in informatics education. Collectively, the papers will raise questions on ways to extend the reach of emerging information ethics research. The session should spark a conversation about the relationship of research and teaching in one of central domain of iSchools.

- Luciano Floridi. (2007). "Information ethics: its nature and scope," In *Moral Philosophy and Information Technology*, edited by Jeroen van den Hoven and John Weckert. Cambridge: Cambridge University Press.

- James H. Moor. (2005). "Why we need better ethics for emerging technologies," *Ethics and Information Technology* 7: 111-119.

- Rafael Capurro. (2006). "Towards an ontological foundation of information ethics." *Ethics and Information Technology* 8: 175-186.

Paper Presentations

1. "Using Immersive Technology to Teach Ethical Dilemmas"

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW) (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW) This paper presents body of interdisciplinary research that is the foundation of a course that is one of the core requirements of a new undergraduate informatics curriculum. The course, Ethics and Information Technology, explores the ethical dilemmas that exist where human beings, information objects, and information systems interact. Modular in design, the course initially is designed to cover interpersonal engagement through online games and virtual environments, maintaining the integrity of digital content in a networked world, and balancing trade offs between secrecy (security) and openness of code, data, and information systems. Students explore the technological underpinnings of associated technology systems, experiment with individual and group interaction with technologies, and examine the mechanics of ethical and unethical behaviors. The course tests the notion that the most effective way to explore how new technologies relate to integrity, truthfulness, trust, respect for privacy and individuality, as well as the variations in ethical behavior across gender, racial and ethnic group, socioeconomic class, sexual orientation, and global cultures is to become immersed in a technological environment where unethical behavior as well as ethical norms can be safely and confidentially tested, evaluated, observed, and experienced.

The paper will review an emerging literature in three areas: (1) the implications of social computing on ethical behavior, the characteristics of the "NetGeneration" regarding the use of new technologies, (2) teaching and learning innovation, and evidence of evolving moral beliefs and behaviors in the online environment, and (3) the challenges presented by specific technologies, including online multiplayer games, image editing, collaborative authoring, and open source coding conventions. The paper will then demonstrate how this base of research literature informs the design and implementation of the course.

2. "Understanding Information Ethics and Policy: Integrating Ethical Reflection and Critical Thinking into Policy Development"

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW) (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW) The development and implementation of information policies must begin with an understanding of underlying values, such as protecting the privacy of personal information, transparency of government information, freedom of expression and protection and use of intellectual property from different cultural perspectives and points of view. As Capurro reminds us: "As a self-referential process ethics is an unending quest on explicit and implicit use of the moral code, that is to say of respect or disrespect, with regard to individual and social communication. . . . There is, indeed, no unbiased ethical observer. (Capurro et al. 2007, 21). In her work on ethical warrant, Berghol (2002) has noted the importance of understanding context. She addresses problems of creating ethically based, globally accessible, and culturally acceptable knowledge representation and organization systems, and foundation principles for the ethical treatment of different cultures. Her concept of "cultural hospitality" provides a theoretical framework for the ethical warrant of knowledge representation and organization systems, and it helps in representing new knowledge and organizations systems and for revising existing systems, and the diverse cultures to which each individual belongs.

Using these two foundations, this paper will provide an overview of some of the critical information ethics issues challenging policy makers in government and in organizations. Using policy frameworks and several models for ethical reflection and critical thinking, the paper will present the framework for a case-based learning approach with a model for critical thinking to present an undergraduate course on Information Ethics and Policy.

- Berghol, C. (2002). "A Proposed Ethical Warrant for Global Knowledge Representation and Organization Systems." *Journal of Documentation* . 58 (5): 507-532

- Capurro, R., and T. Frühbauer. (2007). *Localizing the Internet: Ethical Aspects in Intercultural Perspectives* . Munich: Wilhelm Fink Verlag.

- Paul, R., and L. Elder. (2004). *The Miniature Guide to Understanding the Foundations of Ethical Reasoning* . Dillon Beach: The Foundation for Critical Thinking.

3. "Integrating Social Justice in Undergraduate Informatics Education"

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW) (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW) This paper presents the theoretical and practical issues that confront faculty who introduce social justice in their undergraduate

informatics classes. According to Adams, Bell and Griffin (1997), social justice education includes both experiential pedagogical principles that help students understand the meaning of social difference in real world and virtual social systems, as well as interdisciplinary subject matter that analyzes multiple forms of oppression such as racism, sexism, heterosexism, and classism. The goal of social justice education is to enable students to become conscious of their operating worldview and to be able to examine alternative ways of understanding the world and social relations. Consequently, social justice education has the potential to prepare undergraduate students to become citizens and professionals who are not only able to effectively design, implement and manage information systems and technologies, but who are also (1) sophisticated in their understanding of social interactions amongst diverse groups; (2) able to work democratically with diverse others; and (3) able to critique current social relations and to envision more just and inclusive possibilities for social life in real and virtual contexts.

The paper will be divided into three sections. The first section will lay out the theoretical foundation and framework on which the author's teaching practice is based. The second will consist of the design of course (IST431: The Information Environment) that illustrates this approach. The third section will examine the practical challenges for social justice education – what we should know about ourselves in order to be effective teachers and what we need to know about our students.

- Maurianne Adams, Lee Anne Bell, and Pat Griffin. (1997). *Teaching for Diversity and Social Justice: A Sourcebook*. NY: Routledge.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

Understanding Information Ethics and Policy: Integrating Ethical Reflection
and Critical Thinking into Policy Development

Toni Carbo
Professor, School of Information Sciences and
Graduate School of Public and International Affairs
University of Pittsburgh, 602 IS Building, SIS, 135 N. Bellefield
Pittsburgh PA 15260 USA
412 624-9310
tcarbo@sis.pitt.edu

ABSTRACT

“By Toni Carbo
University of Pittsburgh

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW) (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW) The development and implementation of information policies must begin with an understanding of underlying values, such from different cultural perspectives and points of view. As Capurro reminds us: “As a self-referential process ethics is an unending quest on explicit and implicit use of the moral code, that is to say of respect or disrespect, with regard to individual and social communication. . . . There is, indeed, no unbiased ethical observer. (Capurro et al. 2007, 21). In her work on ethical warrant, Berghol has noted the importance of understanding context (Berghol 2002). Her concept of “cultural hospitality” provides a theoretical framework for the ethical warrant of knowledge representation and organization systems, and it helps in representing new knowledge and organizations systems and for revising existing systems, and the diverse cultures to which each individual belongs. Using these two foundations, this paper provides an overview of some of the critical information ethics issues challenging policy makers in government and in

organizations. Then, building on policy

frameworks and several models for ethical reflection and critical thinking, the paper will describe briefly a case-based learning approach with a model for critical thinking for a proposed undergraduate course on Information Ethics and Policy.

CATEGORIES AND SUBJECT DESCRIPTORS

GENERAL TERMS

KEYWORDS

Information Ethics, Undergraduate Education, Information Policy

1. INTRODUCTION

20.1 1.1 Background and Context

Ethics as reflection on morality is widely accepted among philosopher beginning with Aristotle, the founder of ethics as an academic discipline. As Rafael Capurro, who heads the International Center for Information Ethics (ICIE), reminds us:

As a self-referential process ethics is an unending quest on explicit and implicit use of the moral code, that is to say of respect or disrespect, with regard to individual and social communication. In other words, ethics observes the ways we communicate with each other as moral persons and the ways this moral identify is understood. There is, indeed, no unbiased ethical observer.[1]

In the rapidly changing environments in which information professionals work and in which iSchool educators teach, ethical issues are increasingly being raised, demanding attention and efforts toward resolution. Of particular interest for iSchools and the information field are those related to information throughout its life-cycle. Originally called the ethics of information in society, this topic has more commonly been referred to as Information Ethics (IE) [2]. Beginning in 1997, conferences on the Ethics of Electronic Information in the 21st Century were held at the University Memphis [3]. The subject of IE has also gained international interest as demonstrated by recent activities, starting with the first UNESCO Conference of InfoEthics in 1997 [4]. In 2003, an invitational conference was held in Karlsruhe, Germany under the leadership of the ICIE with support from the Volkswagen-Stiftung [5]; in 2004. The World Summit on the Information Society (WSIS 2003) developed the “Declaration of Principles: Building the Information Society: A Global Challenge in the New Millennium,”[6] and in February 2007, the First African Conference on Information Ethics was held in South Africa (First African Conference on Information Ethics) [7]. Building on the 2007 conference, in August 2008 a UNESCO and the South African government sponsored training workshop on IE and E-Government in Sub-Saharan Africa will be held in South Africa.

Of course, many professional associations in the information field have had codes of ethics for years, and the Association for Library and Information Science Education (ALISE) recently established a taskforce to develop a code of ethics for ALISE, with a report due at its annual conference in January 2009. The topic of Information Ethics has been addressed in the literature, with an annual review of the literature on the topic first appearing in the Annual Review of Information Science and Technology in 1992 [8]. Over the past nearly 20 years, several iSchools have incorporated information ethics into their curricula, both as stand-alone courses and as part of other courses, many of which are required core courses.

2. IE EDUCATION AT PITT'S iSchool

In 1989 the School introduced a lecture series on "The Ethics of Information in Society" to help in shaping the Information Ethics course that was introduced the following year.

The course seeks to provide a background to applied ethics as a prelude to learning the skills of ethical decision-making and, then, to applying those skills to the real and current challenges of the Information Profession. The scope of the coursework and discussions includes decision-making and challenges related to information sources and services in all formats and media; to the Internet and other digital sources (cyber ethics); and to information-related topics in management. The objectives of the course, described as what students will be expected to have at the completion of the course, are:

- Developed a better understanding of themselves (in the ongoing endeavor to "Know thyself");
- Learned how to identify an issue, reflect on it (which is ethics) and make a decision that is moral;
- Engaged in reflective thinking and careful choice of words, which result in civil discourse;
- Developed an understanding of the art and science of applied ethics as related to the main challenges currently confronting the Information Profession.

The course is not a philosophy or religion course, but instead concentrates on the application of ethical reasoning to the Information Profession, with its many, diverse specializations. It is divided into three main sections: an introduction to applied ethics, the necessary steps for facing up to and resolving a moral dilemma and making a decision, and ethical issues in our field. The approach combines "knowing how" with "knowing why" and concentrates on the many questions to be asked in resolving complex issues, beginning with the individual. It is about each of us as an individual even though it is often easier to think of someone else's ethics, and also relates to interactions with other components of life (e.g., the environment, animals, etc.). The importance of the relationship between one individual and another and the need to learn to understand that each of us is a human worthy of respect – that we are fundamentally the same – is stressed. The initial assignment is for students to read and reflect on the U.N. Universal Declaration of Human Rights and Martin Luther King's "I Have a Dream" speech and to write a short paper on the Fundamental Moral Experience. They then identify a personal problem or issue (whether related to personal or professional life) and work throughout the course towards solution of the problem.

For the course, the definition of ethics used is:

- Ethics is the art and science
- that seeks to bring sensitivity
- and method to the discernment of moral values. (Stephen Almagno and Toni Carbo)

Students read a series of books and articles, which change over time and are drawn from a number of disciplines, including philosophy, library and information science, business, and many others. The readings usually include at least two books, such as Stephen Carter's *Integrity* and the Dalai Lama's *Ethics for the New Millennium*, as well as articles from the *Journal of Information Ethics*, the *Harvard Business Review* and numerous other journals in the library and information field and from other disciplines.

3. MODELS FOR ETHICAL DECISION-MAKING

3.1 Context

Several readings and references to websites (including ICIE's website) have been included throughout the course to provide the context for ethical reflection and decision-making, in particular to emphasize the importance of different perspectives from cultural, social, and individual viewpoints. This broader understanding is essential to any course, especially this course. The course is taught in English to English-speaking students and is thus limited in the readings that are used, although students are encouraged to bring in examples from their own readings in different languages and from experience in their own countries and cultures. I have consistently emphasized my own limitations and encouraged others to broaden the views of everyone in the course.

One very helpful, if complex, source for understanding context is Clare Beghtol's work on ethical warrant [9]. Dr. Beghtol, who is on the faculty of the University of Toronto, Faculty of Information Studies, draws upon her extensive research to address problems of creating ethically based, globally accessible, and culturally acceptable knowledge representation and organization systems and foundation principles for the ethical treatment of different cultures. Basing her work on the U.N. Universal Declaration of Human Rights, she presents the concept of "cultural hospitality," to act as a theoretical framework for the ethical warrant of knowledge representation and organization systems. She concludes that the concept of cultural hospitality is promising for assessing the ethical foundations of systems for representing new knowledge and organizations systems and for revising existing systems. Reminding us that each individual belongs to a number of different cultures at different levels (e.g., living in one country, speaking different languages, adhering to policies and practices of different religions and/or political parties, belonging to different social organizations, etc.), she notes that individuals may and do legitimately disagree. She also argues that the boundaries among cultures can themselves be fuzzy and create tensions within an individual. In her discussions of the relationships of any particular culture to its information needs and systems and noting differences among oral and written cultures, she raises questions concerning what kinds of information people need; what they do with it; the extent to which they value it; and whether they choose to perpetuate the information. Her thoughtful, if somewhat densely packed, paper raises several questions, such as whether principles of cultural hospitality can be used to develop culture-neutral systems and theories, which deserve much more discussion. Her paper continues to stimulate thinking and discussion by students.

3.2 Models and Frameworks

3.2.1 The Wheel

Over the fifteen years of teaching the course, we have used a series of models to assist students with ethical reflection and decision-making. Initially, Prof. Almagno used the model of a wheel with four groups of questions in the center hub: 1.) What; 2.) Why? How? Who? When? Where? 3) Foreseeable effects? And 4.) Viable alternatives. The spokes of the wheel are: Creative/imagination, reason/analysis, principles, affectivity, individual experience, group experience, authority, comedy, and tragedy. The questions in the hub provided questions for gathering information and to help in revealing other questions to be asked to determine reality. They also provide a reminder that ethical thinking requires dialogue, even if only with one's self. The spokes serve as evaluation resources through which moral consciousness and awareness can unfold, and they provide a systematic process to address the concerns about how to evaluate each step in making a decision. While some students found the wheel helpful in assisting their decision-making process, many found the steps suggested by the spokes to be somewhat confusing and to overlap.

3.2. Mason et al. Checklist

One example used is that proposed by Mason et al. in Ethics of Information Management [10]. They remind the readers that ethical dialogues are dynamic and nonlinear and suggest a "checklist" of six considerations to be taken into account when resolving an ethical issue [11]. These steps are:

- 1) What are the facts?
- 2) What ethical principles, standards, or norms should be applied?
- 3) Who should decide?
- 4) Who should benefit from the decision?
- 5) How should the decision be made?
- 6) What steps should be taken to prevent this issue from occurring again?

Step one relates to morally relevant considerations and requires both determining pertinent information (understanding the life cycles involved and identification of key decision-making processes) and identifying all the key stakeholders, their values, motivation, and physical history). This step establishes 'what is.' The second step concentrates on 'what ought to be,' and identifies ethical considerations to be applied. The third step seeks to identify who should take necessary actions, how to ensure that all stakeholders are included, and legitimacy and right to make decisions, as well as the ability to affect a resolution. Step four

addresses the various benefits of all the stakeholders and how to balance these, including both short-term and long-run considerations. Step five addresses the method of decision-making, which must be and perceived to be fair and ethical. The final step recognizes that each decision becomes a precedent and seeks to decide what procedures should be used in the future and what decision will be best for the future. Of course, this process must be used within the specific context of the stakeholders making the decision and within the wider societal context.

This framework has been somewhat useful in the course, but many students have found that it does not provide sufficient guidance for them. A key problem is that it starts with gathering facts before reflecting on the questions related to fully understanding what are the initial questions to be asked to help identify exactly what problems and issues need to be addressed. Also, little guidance is provided to address other questions, especially the second one.

3.3 Woodward Framework

The late Diana Woodward, formerly on the faculty of Drexel University, presents a framework for deciding issues in ethics [12]. She discusses advantages and disadvantages of consequentiality and deontology as bases for ethical reasoning in general, and intellectual freedom in general, concluding that a deontological defense of intellectual freedom is “safer” than one on consequential grounds. The article, while helpful for introducing some of the philosophical foundations and theoretical bases for ethical reasoning, does not provide a useful framework for addressing many practical, “real-world” issues.

3.4 O’Boyle

O’Boyle [13] concentrates on the use of the Code of Ethics from the Association for Computing Machinery (ACM), described as deontological because of its enumeration of rights (what is owed by others) and duties (what is owed to others), within a general ethical decision-making process to determine an action. Providing a sound discussion of earlier writings on the ACM Code, he places the Code within a general ethical decision-making process, he differentiates between human faculties: intellect and will. O’Boyle builds on work of Rest and Kohlberg and identifies a six-stage process: 1) Moral perception and personal knowledge of the moral good (recognition that the problem exists); 2) Moral discernment and personal ability to think logically (stating the problem clearly); 3) Moral resolution and personal ability to think analytically (tackling the complexities of the problem to arrive at an individual position); 4) Moral assessment and personal ability to assess one’s freedom (assessment, including being aware of the double-edged sword of new technologies); 5) Moral decision and personal knowledge of one’s duties (decision, including personal duty and obligations); and 6) Moral action and personal willingness to follow one’s intellect (free will used to take action). O’Boyle finds that the ACM Code is helpful with the first three stages, but not with the other three, and that training is needed to apply the Code effectively. He raises two provocative suggestions: that implementing the Code could be advanced by 1) making an individual ethical audit part of an employee’s performance review, and by 2) hiring people who have some understanding and training in ethical behavior. The students have found it helpful in applying codes of ethics to actual situations and problems, and they find the questions to be provocative. Because the article concentrates specifically on the ACM Code, the six-stage framework also has limitations. Also, it does not include extensive discussion of the various needs and perspectives of the various stakeholders, nor does it address many of the complex cultural issues.

3.4 Ottoson

One example from the corporate sector was Gerald Ottoson [14], a retired industry executive, who is now a consultant. Ottoson has conducted numerous ethics seminars over the years for workers in corporations. His approach is to spend a small amount of time on fundamental ethical values (e.g., honesty, mercy, justice, etc.) and to concentrate on models to examine real cases. “A Suggested Pattern of Inquiry,” is presented as a checklist in reviewing and evaluating past actions. The model uses a series of questions: 1) Who should make the decision (includes where the legitimate power to make the decision lies, limits on authority, obligation/responsibility, need for knowledge—the essential facts, and neutrality/objectivity); 2) Who are the stakeholders (principle of regard for others); 3) What are the alternatives (including competing claims, costs/benefits, etc.); and 4) How should the decision be reached (inclusion in the process, perception of involvement, avoidance of paternalism). He notes that decision-making is always a compromise; there is no “perfect” solution; and that there will always be some regrettable aspects of the ultimate decision. He also realistically argues that the final decisions often leave the decision maker feeling a bit uncomfortable. While his framework is incomplete and open to many different interpretations, the students find this example, used in conjunction with others, to be helpful because of its emphasis on the workplace and its realistic questions.

Dr. Christine Altenburger, a retired faculty member from the University of Pittsburgh Graduate School of Public and International Affairs, taught applied ethics for many years. In her teaching, she developed a series of principles and a framework (unpublished), which she has given permission to use in our classes. The basic principles she identifies, summarized from those frequently found in the literature, are: 1) Do no harm. Do good if possible. 2) Observe the cannons of justice. Be fair. 3) Respect the rights, dignity, and freedom of all individuals. She also presents a flow diagram, beginning with gathering facts, leading to analysis and judgment, and incorporating decision loops to reconsider answers to questions.

These and numerous readings have been used throughout our course, but none provided the kind of framework or guide needed for our students as they worked towards resolutions of their problems. One very helpful resource I have used is *The Thinker's Guide to Understanding the Foundations of Ethical Reasoning Critical Thinking Concept and Tools* by Richard Paul and Linda Elder [15]. This brief Guide provides a concise discussion of the importance of critical thinking and the elements of thought, a checklist for reasoning, and a series of questions using these elements. It also summarizes problems of egocentric thinking and introduces a concise and straightforward introduction to the function of ethics, expanding on the discussion of egocentric thought, and addressing problems of "pseudo-ethics". It discusses the differences between ethics and: religion, social conventions, sexual taboos, political ideology, and the law. The authors also remind the readers of why it is important to distinguish among questions of ethics, social conventions, religion and law, and they present a series of elements of ethical reasoning. They propose an eight-step process to determine the logic of an ethical question: 1) purpose (considering an individual's rights and needs as well as those of others); 2) key ethical question(s); 3) information needed to answer the question(s); 4) concepts and principles to guide thinking; 5) main assumptions used; 6) points of view of all stakeholders; 7) main inferences/conclusions (what are the alternatives, are all being considered, etc.); and 8) implications (for self and others, including consequences, questions of harm/good, etc.).

This framework has worked successfully for students to address a wide range of questions and problems. I have also used it effectively as part of a brief introduction to Information Ethics in our required introductory course for Library and Information Science Students, "Understanding Information," in other classes, and in continuing education workshop. In the introductory course, students worked in groups of six to discuss the process they would use (walking through the eight steps) to respond to one of several hypothetical examples based on a real-world case. In each case, the students indicated that the framework encourages them to ask many questions and to examine different perspectives and issues. They also noted the usefulness of the framework in working through individual problems. A significant flaw in the framework is that it focuses on logic and does not recognize the individual's emotions and subjective feelings. This leads to a discussion of the importance of recognizing that no human is ever totally objective and of learning one's own biases, personal values, and cultural perspectives. The Guide seems to be the best tool in our courses to help students work through ethical reflection towards making moral decisions, because it focuses on steps in critical thinking and encourages students to work through the steps, looping back to earlier steps, in the process. The steps rely on logic and objectivity and do not take into account the fact that no human is ever completely objective. This shortcoming can, of course, be addressed by the teacher through readings and by raising questions in the discussion. The guide also serves to help stimulate discussion among students and encourage raising questions about the many options to be considered throughout the steps.

4. PROPOSED UNDERGRADUATE COURSE

At this point, the undergraduate course is in the earliest planning stages. Several challenges must be addressed: 1) teaching undergraduates, rather than graduate students, 2) making the course both attractive and relevant for undergraduates, and 3) incorporating appropriate technologies. Although I have had the occasional undergraduate in my classes (usually from our BSIS program or the Honors College, with one student from the School of Engineering), I have not taught undergraduates. While I have done some research about differences in expectations, learning styles, and experiences in the Social Web, I must admit to some trepidation.

Recognizing that an increasing number of students in my introductory courses enter directly from college and knowing that students in this upper-level course will be predominantly college seniors reduces this concern to a great extent.

Building on the extensive use of case studies and online discussions and group interaction in my courses, I enrolled in a day-long workshop in case-based learning. This workshop was extremely helpful and helped me determine that this was the approach to use. I have decided to use cases based on three primary areas: 1) intellectual property protection and use (particularly

downloading of music and videos); 2) freedom of expression (e.g., restrictions on hate speech at universities, content on Facebook, MySpace, etc.); and 3) gaming. I am currently examining the growing literature on these three areas, especially existing case studies, to determine which readings and examples to include. I plan to use fiction and non-fiction, if possible, getting the assistance of an undergraduate to help identify pertinent literature.

Simulation models related specifically to Information Ethics are starting to appear in the research literature. At the ALISE 2008 conference, I learned of one being developed by Ken Fleischmann at the University of Maryland and his colleagues [16], and we have agreed that we will collaborate. He will make available the beta version when it becomes available from his research project, which will probably not be for at least two years, and I will provide ideas on content and presentation. In addition, as part of a proposal I am developing with colleagues in Information Science and Telecommunications, we will identify ways our two projects (if ours is funded) can share information and learn from each other's. Even if our proposal is not selected to go forward as the single proposal from the University, or if it is not funded, I plan to work with these colleagues to develop and incorporate simulations into the course.

This work is in the earliest stages, and I would appreciate comments and suggestions for the proposed course to help make it as good as it can be.

5. REFERENCES

- [1] Capurro et al. 2007, 21
- [2] Froehlich 1992 and 2004
- [3] See, e.g., Mendina 2000, Rockenbach and Mendina 2003.
- [4] Info-Ethics: First International Congress on Ethical, Legal and Societal Aspects of Digital Information. Sponsored by UNESCO. Monaco, March 10-12, 1997. In: Carbo, T. (Ed.) The International Information and Library Review, 29(2) and 29(3/4).
- [5] International Center for Information Ethics Symposium, Karlsruhe, Germany, 4-6 October 2004. IJIE: International Journal of Information Ethics 2(11) <http://container.zkm.de/ijie/n0002/ijie>
- [6] World Summit on the Information Society. Geneva 2003-Turin 2005. Declaration of Principles. 12 December 2003. (WSIS-03/GENEVA/DOC/4-E)
- [7] First African Information Ethics Conference: Ethical Challenges in the Information Age, Pretoria, South Africa. 5-7 February 2007. (2007) Proceedings. International Review of Information Ethics. number 7. <http://www.i-r-i-e.net>
- [8] Froehlich, T. J. 1992. Ethical considerations of information professionals. Annual Review of Information Science and Technology, 27, 291-324.
- [9] Beghtol, Clare. 2002. A Proposed Ethical Warrant for Global Knowledge Representation and Organization Systems. Journal of Documentation. 58, no. 5 (2002), 507-532.
- [10] Mason, Richard O., Florence M. Mason, and Mary J. Culnan. 1995. Ethics of Information Management. Sage, Thousand Oaks, California. 1995. 85-107.
- [11] Ibid., 103-104
- [12] Woodward, Diana. A Framework for Deciding Issues in Ethics. 1990. Library Trends. 39 (summer & Fall 1990). 8-17.

- [13] O'Boyle, Edward J. 2002. An Ethical Decision-Making Process for Computing Professionals. *Ethics and Information Technology*. 4 (2002) 267-277.
- [14] Ottoson, Gerald H. A Model for Ethical Decision-Making. 1988. *Ethikos*. July/August 1988. 8-11.
- [15] Paul, R., and L. Elder. 2006. *The Thinkers Guide to Understanding the Foundations of Ethical Reasoning*. Dillon Beach: The Foundation for Critical Thinking.
- [16] Fleischmann, K. R., Robbins, R. W. Robbins, and Wallace, W. A. Educational Simulation for Information Ethics. Presentation at the Association for Library and Information Science Education. January, 2008. Philadelphia, PA.

Ethics, Information Technology and
Today's Undergraduate Classroom

Paul Conway
University of Michigan
1085 S. University Ave.

Ann Arbor, MI 48109
734-615-1419
pconway@umich.edu

ABSTRACT

This paper highlights interdisciplinary research grounding a course that is one of the core requirements of a new undergraduate informatics curriculum. Ethics and Information Technology explores the ethical dilemmas that exist where human beings, information objects, and information systems interact. The course tests the notion that the most effective way to explore how new technologies relate to integrity, truthfulness, trust, respect for privacy and individuality is to become immersed in a technological environment where unethical behavior as well as ethical norms can be safely and confidentially tested, evaluated, observed, and experienced. The paper will summarize an emerging literature in three areas: (1) the theories of ethics and information technology, (2) the characteristics of the “Net Generation” regarding the use of new technologies, and (3) the central role played by “trust” in assessing the ethical implications of new technologies, including online multiplayer games, image editing, collaborative authoring, and open source coding conventions. The paper will then demonstrate how this literature informs the design and implementation of the course.

Topics

Information policy, ethics and law

Keywords

Undergraduate education, Information ethics, Pedagogy

21. INTRODUCTION

A new course, Ethics and Information Technology, is a significant contribution by the University of Michigan’s School of Information to a new undergraduate concentration in Informatics. The course is one of four required core courses for the newly approved major, which itself represents a deep collaboration among faculty across three University

schools. The uniqueness of the undertaking raises the stakes, challenging the faculty of each school to define a rich field of study in ways that resonates intellectually across the entire partnership. The nature of this cross-campus collaboration influences the design of new courses, as well as the

approaches to undergraduate education. Ethics and Information Technology tests the notion that the most effective way to explore how emerging technologies relate to information ethics is to immerse students in a technological environment where unethical behavior as well as ethical norms can be safely and confidentially tested, evaluated, observed, and experienced. In this regard, the course exists at the intersection of technology, ethics, and pedagogy. This paper frames the principal theoretical issues that underlie the design of the course and outlines its most salient pedagogical features.

22. CONTEXT

“Ethics” is variously defined [O.E.D.] as a branch of philosophy concerned with evaluating human action, the study of individual or group values, or a system of defining right and wrong behaviors. When applied to the professions, ethics defines a code of standards governing fair and responsible conduct with other members of a profession and the general public. In association with computer and information technologies, ethics concerns the relationship of systems with the people who use them. [14] Most recently, the concept of information ethics has extended philosophical consideration well beyond the human behavior to explore the ethical properties of information objects. [9]

The study of ethics within the context of information technology is international in scope. The International Center for Information Ethics (ICIE) identifies 104 individuals in over 90 organizations around the world whose primary field of scholarship and teaching is information ethics, with Germany, Japan, the United Kingdom, and the United States having a particularly strong presence in the field. [13] For the United States, ICIE’s selective database lists 32 individuals in 30 universities with a primary focus on information ethics research. Nearly a dozen scholarly journals, continuing sequences of international conferences, and highly touted monographs and compendiums attest to the deepening scholarly interest in ethics and emerging technologies. The teaching of ethics has traditionally been an important element of the university curriculum. The University of Michigan’s Ethics in Public Life Initiative, for example, has compiled a current list of over 300 regularly-scheduled courses that involve ethics as a substantial component. [22] Academic disciplines offering these courses range from the traditional and obvious (e.g., philosophy) to the more subtle and nuanced (e.g., environmental studies). Every professional school at Michigan offers at least one course at the graduate level focused on either professional ethics or ethics policies appropriate to the profession. Those professional schools that offer undergraduate courses or degrees include, but do

not necessarily require the study of ethical issues to complete a major.

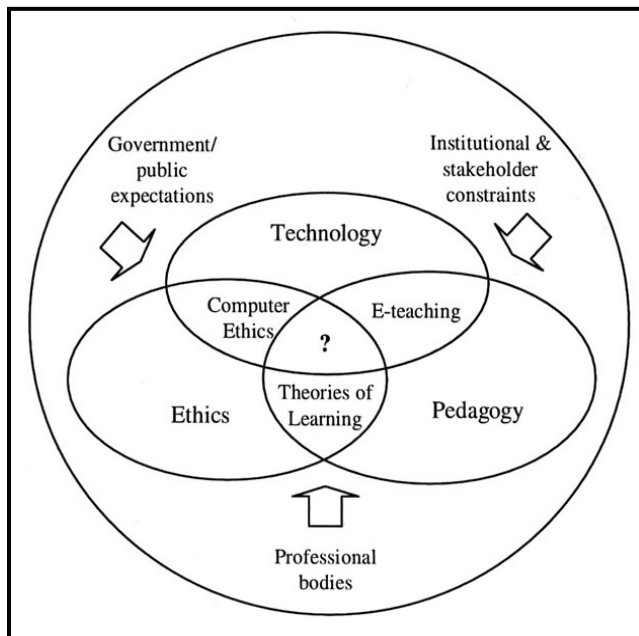
Ethics in the iSchools, however, presents a paradox. Ethics is simultaneously a core principle governing the formulation of their teaching missions and a minor or peripheral area of research. On the one hand, the websites for all of the 19 iSchools describe ethics as one of their important areas of concern. Nearly all iSchools offer specific graduate-level courses that include an investigation of either professional ethics or ethics policy issues. And yet, six of the 19 members of the consortium of iSchools account for a total of only eight faculty in the ICIE database of research specialists in information ethics. Although the ICIE database does not reflect the efforts of faculty who primarily or exclusively teach ethics, it is clear that research on ethics is not at this point in time a central research or teaching strength of most iSchools.

23. A FRAMEWORK

The undergraduate course Ethics and Information Technology exists at the intersection of three distinctive concepts: (1) the philosophy of ethics, (2) the social characteristics of emerging information technologies, and (3) the learning styles of a particular cohort of students. Although literature on the theoretical foundation for each of these areas is vast, there is a particular

Figure 1. The links between technology, pedagogy and ethics

absence of literature that addresses the theoretical



foundations challenges at the intersection. McRobb, Jefferies and Stahl [16] propose a framework that describes research findings in the areas where the concepts overlap and establishes a domain where the concepts of ethics, technology, and pedagogy are most dynamic. Figure 1 is the authors' graphic illustration of their framework.

The McRobb framework and its supporting literature is oriented toward the ethical issues involved in distance learning, drawing for support on the literature of computer supported collaborative learning. The conceptual model represented by the McRobb illustration, however, provides a convenient mechanism for highlighting how the issues that are at the heart of the Michigan course overlap and intersect. The framework includes three pressures that constrain the content of the three issue areas, among them the expectations of a wider audience regarding the evolving role of ethics education in the academy, the limitations imposed by university technology systems, and perspectives of various professional bodies or academic disciplines on the appropriate approaches to ethics education. Each of these constraining issues is at work in the design of a course that purports to exist outside the boundaries of the existing curriculum. Ethics and Information Technology is not only a new course but one that straddles and attempts to incorporate the perspectives of the three disciplines that are contributing to the new Informatics curriculum.

The framework is most relevant in identifying three distinctive intellectual spaces where conceptual overlap creates interesting new perspectives. First, the boundary where the study of ethics and the characteristics of emerging technology overlap, labeled in the figure as "Computer Ethics," provides a set of useful ethical models that are the intellectual structure of course. Second, the boundary where ethics and pedagogy overlap orients the course's interactive design and the flow of the individual course modules. Third, the boundary where technology and pedagogy overlap provides the rationale for an immersive technology environment in which students and instructor model the challenges posed by a suite of technologies.

24. ETHICS AND TECHNOLOGY

Scholarship on the relationships between information technology and ethical beliefs and behaviors have been debated and refined for the past fifty years. How this scholarship informs the teaching of these relationships is less well understood.

Terrell Bynum [1] credits the American philosopher/scientist Howard Weiner (the founder of the science of cybernetics) with foreseeing the enormous ethical and social impacts of information technology and laying the groundwork for the study of computer ethics. Writing in the 1950s, Wiener grounded his ethical theories of computer technology in the view that human beings are complex information feedback system that govern their relationships with other humans and the world around them. Drawing on Aristotle and flowing through Weiner's systems perspective, Bynum proposes that the purpose of human life is to flourish as a person [2] and to do so through a diversity of information processing activities. In this regard, the principal value of information technology is to extend human potential by reinforcing the principles of freedom, equality, and benevolence. Weiner simultaneously defined the ethical underpinnings of Vannevar Bush's Memex and anticipated the collaborative social technologies of

Douglas Englebart. Bynum and others see the culmination of Wiener's vision of "flourishing ethics" in the theoretical constructs of Luciano Floridi's "Infosphere." [8]

James Moor is a bridge from the discipline of computer ethics to the broader world of information ethics. Moor [18] initially defined computer ethics in terms of a policy vacuum that occurs when "new technological capabilities provide new choices for action in an environment where existing policies seem inadequate." Computer ethics includes consideration of both personal and social policies for the ethical use of computer technology. Writing twenty years later, following the explosion of the World Wide Web and the emergence of widespread social computing applications, Moor called for a more flexible and agile approach to investigating the relationship of new technologies to human ethical behavior. Moor [19] establishes a three part progressive model of technology development that relates the maturity of revolutionary technology to increasing ethical complexity. According to his model more people will be involved, more technology will be used, and hence more policy vacuums and conceptual muddles will arise as the revolution advances." In the case of emergent (immature or experimental) technologies, such as the socially oriented tools and systems typified by the over-used moniker Web 2.0, Moor postulates his own "Law," which states that "as technological revolutions increase their social impact, ethical problems increase," because revolutionary technology provides many new opportunities for action "for which well thought out ethical policies will not have been developed." Moor lays out three approaches to study new technologies from an ethical perspective, including investigating new technologies before they have stabilized, using multi-disciplinary approaches to research, and adopting sophisticated ethical analyses to avoid the tendency to revert to simplistic cost/benefit analyses that translate ethical choices into monetary terms. "We need to learn about the technology as it is developing and to project and assess possible consequences of its various applications." The Michigan course places students into a policy-weak environment populated by technologies whose use and abuse are not fully understood.

An important open question at the intersection of ethics and technology is whether the social dynamics of new technologies are generating new ethical models of behavior. In addressing this question, Himma [11] focuses on the role of ethics in informed decision making. He reviews and ultimately dismisses claims that computer ethics has a claim to theoretical uniqueness. "Understanding computing technologies will help to produce well-informed ethical views – regardless of how we characterize those technologies." Rafael Capurro [4] adds another dimension by highlighting the nature of the content that is embedded in or made accessible by networked technologies. He argues for a holistic view of information that is attentive to the mass transformation/transition of content from analog to digital. "In this broader sense information ethics deals with questions of digitalization, i.e., the reconstruction of all possible phenomena in the world as digital information and the

problems caused by their exchange, combination and utilization." Capurro makes an essential connection between communication technologies and the human propensity to share and preserve. A basic moral principle of the information environment, he claims, "is to share knowledge, or the right to communicate in a digital environment which includes the right to preserve what we communicate for future generations." Capurro reminds us that the appropriation of modern information technology is not just a technical also but a culturally-bounded endeavor. Luciano Floridi [9] presents, perhaps, the most well developed philosophical perspective on the ethical issues associated with information and communication technologies. Floridi's "Infosphere" encompasses not only cyberspace but also off-line and analog information spaces. Adopting an object-oriented approach to the design of a new ethical model, Floridi defines moral action as a "dynamic system" arising out of the interaction of seven principal components: 1) the agent, 2) the patient, 3) their interactions, 4) the agent's general frame of information, 5) the factual information concerning the situation that is at least partly available to the agent, 6) the general environment in which the agent and patient are located, and 7) the specific situation in which the interaction occurs. Drawing deeply on environmental physics, Floridi restates that the fundamental principles (or rules) of this dynamic system are grounded in the notion of information entropy – that is the destruction, pollution and depletion of information objects – ought not to be caused, ought to be prevented, ought to be removed, and ought to be protected, extended, improved, enriched and enhanced. [7] Information ecology as a parallel. Floridi's model and the norms it proposes structure the flow of the Michigan course.

The real question for the intersection of ethics and technology is the extent to which new technologies and the ways that people use them foster new rules of ethical behavior (perhaps culturally determined) or whether long standing principles are transferred to new technological contexts. The pedagogical focus of the course is designed in part to explore this issue dynamically and interactively.

25. NET GENERATION PEDAGOGY

A course that examines ethical issues associated with new technologies must necessarily take account of the learning styles and the general attitudes of the undergraduates who enroll. The conceptual design of the course, as well as its intellectual flow of the individual modules and use of technology tools in and outside the classroom are in part predicated on the notion that today's generation of undergraduates is somehow different than previous generations.

Allowing for fluidity in the boundary lines, a case can be made that at any point in time, generational cohorts bound by shared experiences and history exhibit behavioral and attitudinal cohesion. Commentators have labeled the generation of students born since 1981 (a somewhat arbitrary

point in time) as the Net Generation, in part because they have grown up with the widespread availability of personal computers and the Internet. Strauss and Howe [21] characterize the members of the Net Generation as sheltered and protected but pressured to excel, endowed with a strong sense of their own specialness (indicated by high self-esteem), confident, and optimistic. They may be more team oriented than previous generations, more comfortable than average with multi-tasking, and very literate in the realm of digital and visual technologies. Gibbons [10] argues that the affinity of today's undergraduates for information technology "translates into new and different expectation about how to gather, work with, translate, and share information."

The apparent naturalness with which undergraduates embrace new technologies leads some commentators to see technological determinism at work. Lippincott [15] argues that "digital natives" have acquired styles of learning and modes of interpersonal interaction as a direct result of the availability of network technologies, suggesting that the implications of this fact extend beyond the classroom to challenge and change the mission and purposes of libraries and other information services. Nye [20] and most scholars of the history of science and technology reject the deterministic thesis, arguing instead that "people become enmeshed in a web of technical choices made for them by their ancestors," helping to explain why people may seem trapped by the choices others have made.

The design of the Michigan course rejects a deterministic view of the technologies it utilizes but starts with the premise that there are generational differences in the perspectives of students and teachers. Such differences may be manifested genuinely in varying comfort levels with new and emerging technologies. But the course's structure, flow, and assessment methodologies leave open the question of whether Net Generation learners harbor a distinctive ethical world view in the use of these technologies.

26. THE CENTRALITY OF TRUST

The McRobb et al. framework (Figure 1) places a question mark at the center of its analytical framework where ethics, technology, and pedagogy meet. In their review of the associated research literature, the authors found that "there seems to be little interest in the exploration of the intersection of those areas...What appears to be lacking is a good overview of the relationships of the different issues involved." [16]

In adapting the McRobb framework, course designers used the concept of "trust" as an organizing principle for the individual modules. The O.E.D. defines trust as: "confidence in or reliance on some quality or attribute of a person or thing, or the truth of a statement." In placing trust at the intersection of ethics, information technology, and pedagogy, the Michigan course opens students to an examination of trust from multiple perspectives, for example trust in individual identity, trust of the integrity of digital content,

trust in the transparency of open source code and the community of people who create it. As the case studies in Hutchings [12] demonstrate, the notion of trust also enters into the dynamic engagements between teacher and student, as well as student to student interactions inside and outside of class.

In terms of identity with the context of the course, Buchanan and Ess [3] demonstrate that trust bears on the extent to which behavior within a virtual environment is conditioned by knowledge of the identity of the agents within the environment. The creation, modification, and behaviors of avatars in an online environment is a particularly apt example of identity trust. Chesney [5] follows the same line of reasoning in examining the trust metrics underlying such collaborative writing efforts as Wikipedia.

In terms of content, trust is intimately related to the notion of "integrity," which Duranti [6] defines universally for textual documents in terms of reliability and authenticity. "It is generally accepted by all literate civilizations that documents are trustworthy (that is, reliable) because of their completeness and controlled procedure of creation, and which are guaranteed to be intact and what they purport to be (that is, authentic) by controlled procedures of transmission and preservation, can be presumed to be truthful (that is genuine) as to their content." Mitchell [17] reaches a less technical but no less measurable conclusion in reference to trust in visual content. "If an image follows the conventions of photography and seems internally coherent, if the visual evidence that it presents support the caption, and if we can confirm that this visual evidence is consistent with other things that we accept as knowledge within the framework of the relevant discourse, then we feel justified in the attitude that seeing is believing."

27. DESIGNING A COURSE

As conceived at Michigan, Ethics and Information Technology is one of four courses of a new undergraduate concentration in Informatics. The multidisciplinary concentration is a collaborative undertaking among faculty from the university's Computer Science & Engineering division within the College of Engineering, the Department of Statistics within the College of Literature, Science and the Arts (LSA), and the School of Information. The undergraduate concentration is based in LSA but draws on faculty from all three schools. Program development is led by a steering committee of two senior faculty from each of the participating schools. Ethics and Information Technology is guided by the collaborative spirit of the faculty development team, but is being designed and will be taught by a team of two faculty, one from the College of Engineering and one from the School of Information.

Ethics and Information Technology explores the ethical dilemmas that exist where human beings, information objects, and information systems interact. Modular in design, the course introduces students to a variety of ethical models from historical and cross-cultural perspectives and then explores the relevance of these models to a variety of new and

emerging technologies that are inherently social in their construction and use. Initial examples of issues that the course covers in discrete modules include:

- interpersonal engagement through online games and virtual environments,
- the integrity of digital content in a networked world, and
- tradeoffs between security and openness of code, data, and information systems.

Students explore the technological underpinnings of associated technology systems, experiment with individual and group interaction with technologies, and examine the mechanics of ethical and unethical behaviors.

The course has two major instructional objectives: (1) integrate opportunities for direct hands-on technical experiences, in order to enhance understanding of ethical challenges presented by new information technologies, and (2) offer opportunities to participate via in-class discussions, short posts submitted to an online discussion board, and longer written assignments; in order to encourage multimodal contributions by students. Associated with the objective are three related learning outcomes: (1) demonstrate knowledge of current theories in information ethics; (2) apply ethics theories to interpret behavior when using a variety of information technology tools; and (3) evaluate the nature of ethical choices made by self and others when serving various roles.

New information technologies raise knotty issues regarding integrity, truthfulness, trust, respect for privacy and individuality, as well as the variations in ethical behavior across gender, racial and ethnic group, socioeconomic class, sexual orientation, and global cultures. The course tests the notion that the most effective way to explore these issues is to become immersed in a technological environment where unethical behavior as well as ethical norms can be safely and confidentially tested, evaluated, observed, and experienced. The course will mix experiential learning with individual and group interaction with a variety of technologies.

This course is unique in its construction and in its mix of technological tools for instruction. Although a vibrant literature is emerging on approaches to teaching ethics and information technology and on the use of technology in the classroom, relatively little is known about learning processes and learning outcomes in the combination of ethics and emerging social technologies. The course will, in part, form a test environment for learning about how to use innovative technological tools to teach about the ethical dilemmas posed by these same technologies. Significant instruction takes place with the aid of technological tools available through an online virtual environment build on the Sakai platform that most University of Michigan students recognize as CTools.

27.1 Example: Module Assignments

Assignment 1: Virtual Environments: Anonymity

Students will create an avatar in a virtual world and adopt an anonymous persona. They will interact with their classmates

in a secure environment over a period of several days, and then attempt to guess which student each avatar represents. Students will maintain awareness of such factors as whether they encountered any challenges in maintaining anonymity on a long-term basis; whether they employed confounding strategies to actively deter detection; how frustrated they were by others' anonymity; whether their anonymous status conferred a sense of freedom to behave transgressively; the extent to which their or others' personalities were identifiable through language, appearance, or gestures, etc. What are some of the unexpected disadvantages of anonymity in this environment? Apply ethical theories discussed in class in your analysis.

Assignment 2: Virtual Environments: Cheating

Students play videogames with colleagues: (1) cooperatively as a member of a team, and (2) competitively in an individual PvP exercise. (1) Students on each team first must discuss and decide as a group whether their team will collectively condone and engage in cheating behaviors in order to advance their progress, and note the results generated by their decision whether or not to cheat. Each team also must determine whether they have identified evidence of cheating by their opponents, and if so, what their response will be in return. (2) As individual players, students will be encouraged to cheat in one-on-one interactions in order to succeed. They must note their emotions, as well as their intellectual rationalizations, when engaging in transgressive behavior. Is the game more or less enjoyable when both players are known to be cheating? Apply ethical theories discussed in class in your analysis.

Assignment 3: Information Integrity: Critical judgment

In a three-part exercise, students examine a set of photographic images to which various enhancement algorithms and editorial techniques have been applied. Working in teams of two to four individuals, students will then categorize a set of existing digitized photo images according to a four part rendering scheme and evaluate the truthfulness of the images within the scheme. Finally, students will assess how meaningful and trustworthy are a set of images in an online exhibit by comparing and contrasting the messages of the image with the messages of the accompanying text.

Assignment 4: Information Integrity: Collaborative editing roll playing

Students will be assigned randomly to one of four roles: known author; anonymous author; known editor; anonymous editor. Students will play their roles in creating and editing content in a class-limited wiki. Topics chosen for creation and editing will be determined by the class. Students will be given explicit assignments regarding the submission of truthful and untruthful content and on the "rules of engagement" for adding, editing, and deleting content. Students will log their activities and seek to create together a set of trustworthy wiki entries. Discussion during and after the exercise will explore the challenges of establishing and keeping trust.

Assignment 5: Secrecy and Openness: Cryptography

The purpose of assignment is to assess the level of knowledge of the cryptography section of this course. An exam on the topic will be offered. The exam key will be posted on the CTools site in advance of the students taking the exam. However, the key will be posted in an encrypted form. The students will not be told which encryption scheme was used for encoding. Students will have the choice of studying for the exam in a traditional way, or putting their effort towards decrypting the answer key and guaranteeing full credit on the exam.

Assignment 6: Secrecy and Openness: Role playing

Homework is a role-playing assignment. Students are given a list of 'artifacts' of varying levels of importance to them personally, and to society in general. Students will then be asked (individually or in groups) to prioritize artifacts in terms of importance and need for secrecy. The initial evaluation will be from their current perspective as a UM student. Next, students will be randomly assigned a role for role-play. Students will again be asked to prioritize artifacts and reflect upon any changes in their prioritization scheme.

28. CONCLUSION

The design and implementation of an innovative course on ethics and information technology is a somewhat risky undertaking. First, we do not yet know of the market for the concentration and the extent to which the new course will both attract students to the concentration and fulfill critical learning objectives for the new Informatics Program. Second, we cannot anticipate the extent to which experiential teaching in the domain of ethics and information technology will lead to specific learning outcomes. Toward this end, the course instructors have partnered with the university's Center for Research on Learning and Teaching to develop assessment metrics new student feedback mechanisms. Third, we are not sure that the course will or should result in behavioral changes in the students who complete the course. We believe the risk of failure (measured by either low enrollment or low student evaluations) is offset by the opportunity to create a learning environment that serves as a model laboratory for new research on teaching ethics at the undergraduate level.

29. ACKNOWLEDGMENTS

The author offers thanks to David Chesney, College of Engineering, and Marianne Ryan, School of Information, University of Michigan, for collaborating in the development of the course described in this paper. At the University of Michigan, the Provost's Multidisciplinary Learning and Team Teaching initiative and the President's Ethics in Public Life initiative provided generous support for planning and start up of the course.

30. REFERENCES

- [1] Bynum, T. W. 2004. Ethical challenges to citizens of 'The Automatic Age': Norbert Weiner on the information society. *Journal of Information, Communication and Ethics in Society* 2, 2 65-74.
- [2] Bynum, T. W. 2006. Flourishing ethics. *Ethics and Information Technology*. 8, 2, 157-173. DOI=10.1007/s10676-006-9107-0
- [3] Buchanan, E. and Ess, C. 2005. The ethics of e-games. *International Review of Information Ethics* 4 (December), 2-6.
- [4] Capurro, R. 2006. Towards an ontological foundation of information ethics. *Ethics and Information Technology* 8, 2, 175-186. DOI=10.1007/s10676-006-9108-0
- [5] Chesney, T. 2007. An empirical examination of Wikipedia's credibility. *First Monday* 11 (November). http://www.firstmonday.org/issues/issue11_11/chesney/index.html
- [6] Duranti, L. 1995. Reliability and authenticity: the concepts and their implications. *Archivaria* 39, p. 8.
- [7] Floridi, L. 1999. Information ethics: on the philosophical foundations of computer ethics. *Ethics and Information Technology* 1, 1, 37-56.
- [8] Floridi, L. 2002. On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology* 4, 4, 287-304.
- [9] Floridi, L. 2008. Information ethics: its nature and scope, In *Moral Philosophy and Information Technology*, J. van den Hoven and J. Weckert (eds.). Cambridge University Press.
- [10] Gibbons, S. 2007. The academic library and the net gen student: making the connections. Chicago: Amer. Lib. Assoc..
- [11] Himma, K. E. 2007. Foundational issues in information ethics. *Library Hi Tech* 25, 1, 79-94. DOI=10.1108/07378830710735876
- [12] Hutchings, P. (ed.). 2002. Ethics of inquiry: issues in the scholarship of teaching and learning. Menlo Park: Carnegie Publications.
- [13] International Center for Information Ethics. <http://icie.zkm.de/>
- [14] Introna, L. 2005. Phenomenological approaches to ethics and information technology. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/ethics-it-phenomenology/>
- [15] Lippincott, J. 2005. Net generation students and libraries. In *Educating the net generation*. Boulder: EDUCAUSE. <http://www.educause.edu/ir/library/pdf/pub7101m.pdf>
- [16] McRobb, S., Jefferies, P., Stahl, B. C. 2007. Exploring the relationship between pedagogy, ethics and technology: building a framework for strategy development. *Technology, Pedagogy & Education* 16, 1, 111-126. DOI=10.1080/14759390601168122
- [17] Mitchell, W. J. 2001. The reconfigured eye: visual truth in the post-photographic era. Cambridge: MIT Press, p 43.
- [18] Moor, J. H. 1985. What is computer ethics? In T. W. Bynum, ed. *Computers and Ethics*. Basil Blackwell, pp. 266-275.
- [19] Moor, J. H. 2005. Why we need better ethics for emerging technologies. *Ethics and Information*

Technology 7,2, 111-119. DOI=10.1007/s10676-006-0008-0

- [20] Nye, D. E. 2006 [check pub date]. Technology matters: questions to live with. Cambridge: MIT Press.
- [21] Strauss, W. and Howe, N. 2006. Millennials and the pop culture. Great Falls, VA: Life Course Assoc.
- [22] University of Michigan, Ethics in Public Life Initiative.
<http://www.umich.edu/pres/ethics/about/>

Ethics and Social Justice in Undergraduate Informatics Education

Lynette Kvasny
 Pennsylvania State University
 329C IST Building
 University Park, PA 16802
 814-865-6458
 lkvasny@ist.psu.edu

ABSTRACT

This paper examines social justice as a praxis oriented approach for infusing ethics and multiculturalism in undergraduate informatics education. Social Justice education includes an interdisciplinary study of information and communication technologies (ICT) that analyzes unequal power relations that exist in society, and a set of interactive, experiential pedagogical principles that help students understand the meaning of social difference in their personal lives, in their professional lives, and in their civic lives. The paper summarizes the theoretical basis of social justice and its relationship to ethics, multiculturalism and ICT. The paper also presents a pedagogical approach used in the design and implement of an existing undergraduate informatics course. Social justice education has the potential to prepare students who are (1) able to critically evaluate the intended and unintended consequences of ICT adoption and use, (2) sophisticated in their understanding of social institutions in shaping the meaning and intended uses of ICT, and (3) able to work with diverse others.

Topics

Information policy, ethics and law

Keywords

undergraduate education, ethics, social justice, pedagogy

31. INTRODUCTION

ICTs provide new and exciting ways for people to work, coordinate schedules, share information, maintain relationships and seek entertainment. In this paper, the information and the technologies that are employed by humans to create, organize, transfer and utilize that information in a networked environment are collectively known as "the information environment". This information

Copyright and Disclaimer Information

environment is changing the ways in which we interact,

communicate, and function on the job and in our daily lives. Emerging technologies also raise new economic, cultural, legal, ethical, and social issues that are of grave importance to society.

The College of Information Sciences and Technology (IST) at the Pennsylvania State University offers an existing course, IST431: The Information Environment, which surveys these diverse issues. This course examines how and why the evolving information environment influences the design, development and use of ICT. This course also examines the social impacts of ICT on people, organizations, and societies. The information environment is a messy. The issues are not clearly identified, there are competing interests, and solutions are extremely difficult to develop because there is no one, best answer to the problem. To assist students in developing the skills necessary for coping with this situation, the course employs a social justice approach to education. The subject matter of the course is learned through a series of problems that enable students to learn about the layers of the information environment and the factors embedded in them. The learning format emphasizes a student-centered classroom activities and team-based projects that feature participation, inclusiveness, affirmation of human agency, social responsibility, and collaborative work in small groups. Through individual and group learning activities, students and the educator examine the influence of environmental factors on ICT and the impact of ICT on the environment. This paper summarizes the primary theoretical and pedagogical features of this course and the broader implications of social justice pedagogies for educators.

32. THEORETICAL FOUNDATIONS

According to Adams, Bell and Giffin [1], social justice education seeks the full and equal participation of all groups in a society. This is a vision of society in which diverse social groups mutually shaped ICT to meet their needs. In such a society, individuals are both self-determining (able to develop their full capabilities) and interdependent (capable of interacting democratically with others). Students should have both a sense of social responsibility as well as a sense of their own agency to speak and act in a manner that supports their beliefs.

Social justice education is also praxis oriented [2]. Praxis posits that theory and practice are interwoven, with practice being shaped by informal and formal, tacit and explicit

theories. Hence, the way that we identify and articulate problems that exist in the information environment, the solutions that we entertain as viable, and the methods that we choose as appropriate for realizing those solutions are all theoretical as well as practical questions [1].

Consequently, theory serves three important functions. First, theory enables us to reflect critically upon our intentions, decision-making, and the means we use to actualize our choices in the design of ICT. Computerized information systems, for instance, are generally designed for some idealized audience. Ethical problems surface when users with different knowledge requirements attempt to use the system. Hoesle [3] notes that computerized information system use requires people to act and think in prescribed ways that privilege Western cultural traditions while marginalizing the cultural traditions of others. Third-world cultures have unique forms of knowledge embodied in cultural artifacts such as mythologies, storytelling, proverbs, art and dance. While these forms of knowledge are legitimate and deserve to be preserved, Hoesle [3] posits that they cannot be equally well-represented through computerized ways of storing, representing and transmitting knowledge. In the 15 years since Hoesle authored this manuscript, ICT has advanced significantly. With Web 2.0 technologies, relatively low skilled users can author content in a variety of social media formats such as video, podcasts, wikis, and blogs. With these diverse formats, users now have more options for representing their unique worldview, and Hoesle's concern may be less salient. The challenge may be less a problem of inadequacies in computerized representations of knowledge, and more a problem of providing access to the tools and competencies that make self authoring possible. Open source applications provide lower cost technology solutions to organizations and individuals seeking alternatives to proprietary information systems. Both Web 2.0 and open source are based on community-oriented design intentions and choices, and expand the usage options for non-traditional users of ICT.

Second, theory has the potential to help us stay conscious of our position as historical subjects. Ethical implications arise from these historically situated power relations. For instance, ICT is often seen as an essential component of development and democratic practices. This viewpoint is expressed in the context of developing countries and historically underserved groups in the developed world. These groups are seen as excluded from the information society, and, consequently, should apply ICT or risk being left behind. This line of reasoning not only supports the technology imperative, but also subscribes to a development ideology that is based on a concept of history as linear and universal. All societies and people go through the same stages until they reach the same economic levels as countries and people considered developed [4]. The assumption is that ICT is neutral and nondiscriminatory in that it potentially provides equal opportunities to everyone; equal access to ICT ensures equal benefits. However, this ignores historical power relations such as colonialism, racism, and sexism to which ICT is

inextricably linked. The perception that ICTs are necessary for development also ignores that technology invents its own needs such as software and hardware upgrades and training [5], as well as problems such as the environmental degradation that results from a Western lifestyle. Ecologically, it is impossible for the whole world to adopt a Western lifestyle [3]. Yet a Western lifestyle is offered as the standard for development.

Third, and perhaps most importantly, theory provides a framework for interrogating our practices and creating novel approaches to addressing issues in the information environment. Many of these issues, such as social and digital exclusion [6][7][8], privacy [9][10][11], and community and democracy [12][13], have strong ethical and multicultural implications. For instance, the digital divide is at least as much about cultural differences as it may be about disparities in income and infrastructure. Because the internet enables users to focus solely on receiving information from and communicating with like-minded individuals, Paterson [4] worries that the internet may foster fragmented existences that exacerbate existing social differences based on ethnicity, gender, geography, age and race.

The central implication underlying these issues of ethics and multiculturalism is how diverse people in different parts of the world experience the effects of the computerization. For some, the challenge is to continuously innovate and create new technologies, others grapple to keep up with the continuous changes enabled by ICT, and still others struggle to put in place the basic infrastructure that may allow them to participate in the global, knowledge-based society.

Praxis forces us to examine how power operates through normalizing discourses about ICT and power relations. Praxis reminds us to continually ask "in whose interest" prevailing systems operate [1]. The question of who is served by ICT provides a useful analytic tool for exposing hierarchical relationships, as well as the hidden advantages and penalties embedded in technologies that are purportedly neutral. A commitment to social justice requires an ethical attitude towards inclusion of diverse cultures and the potential of ICT to improve the capacity of human agents to transform their world.

33. COURSE DESIGN

The social justice approach evolved over five years of teaching and continuously improving The Information Environment course. The context in which the course is situated as well as the pedagogical mechanisms that were used in the evolution of the course are presented.

33.1 Context

The College of IST prepares undergraduate students to confront and manage the complex interactions among technology, people, and information. Students are broadly trained in the areas of software engineering and networking, user interface design, information systems development, human-computer interaction, and social informatics. Our

curriculum integrates these areas through problem-based learning, often with industry-based collaborators. Students emerge from our program with a systems view of ICT, experience working in teams on unstructured problems, and an understanding of the contextual factors (from the technology itself to cognition, ethics, economics, and policy) that shape the design, adoption and use of ICT. In 2007, over 90% of our graduates were placed in technology-oriented career positions in diverse industries such as consulting, banking, manufacturing, healthcare and government. Given the industry orientation of the IST curriculum and my professional experience and graduate education in information systems at an American business school, the initial design of the course focused primarily on professional ethics. Using the ACM and IEEE Codes of Professional Ethics as a basis, students were grouped in teams and presented with fictitious scenarios that portrayed ethical dilemmas that they may face in their careers. The scenarios included breaches in the confidentiality of information, programmers' and analysts' obligation to act in a socially responsible manner, use of company computers for personal benefit, ownership of intellectual property, and electronic monitoring of workers. While the scenarios generated good discussion, students didn't seem to internalize the material. A second approach was to replace the fictitious scenarios with real-world cases from the popular press. Student teams were asked to evaluate and debate the behaviors of the various stakeholders presented in the cases. Again, most students failed to personalize the ethical issues. To obtain insights as to how I could improve the course, I instituted a student quality team. In our college, student quality teams assess the classroom learning environment and provide educators with feedback to help them improve course design. The student quality team program is student-organized and managed. Team leaders are trained in quality principles, team management, survey design, and diplomacy. They also receive nominal salaries. Team members are recruited from students enrolled in the course, and they receive one credit of independent study for their effort. Students noted that the scenarios seemed abstract and removed from their daily lived experiences. They had little formal experience working in a business organization. Consequently, they had yet to develop a strong professional identity. Also, because ethics was isolated in a course module and presented solely as a professional concern, the importance of ethical considerations was minimized. While considering the student feedback, I attended the Multicultural Teaching Academy at the Schreyer Institute for Teaching Excellence at the Pennsylvania State University. Social justice was presented as an approach for infusing multiculturalism and ethics in undergraduate education. Social justice also fit with my research interests in ICT and social exclusion, and personal commitment to improving the recruitment, retention and academic excellence for women, students of color and international students who are severely under-represented in our program. Hence, social justice formed the basis for the third iteration.

33.2 Pedagogical Approach

A pedagogy respectful of social justice and human diversity begins with the educator's sense of social responsibility [14]. Education is fundamentally ethical and political in nature. Ethics and politics are inherent in the teacher-student relationship, in the physical conditions of the classroom, in the grading, and in the readings and course content that are chosen or excluded from the syllabus.

To achieve a democratic learning environment that is inclusive and meaningful to students, the first day of class begins with the educator presenting an overview of the course and the key learning objectives. These learning objectives include:

- identify major themes in information policy studies (e.g. ethics, privacy, workplace monitoring, security) and relate these themes to the applications of ICTs
- use conceptual frameworks to frame analysis of ICTs in their context of use, at varying levels of analysis (e.g. individual, organizational, societal, global) reflected in the information environment
- analyze the impacts of information systems and technology, and make appropriate policy recommendations

Students are then broken into teams and asked to identify three important topics that they would like to explore. The class then organizes the topics into themes that are then incorporated into the syllabus. Table 1 provides an example of topics that were explored during the Spring 2007 semester.

Modules	Sample Topics
Theories of technology and social change	technology determinism, social determinism, social informatics
Technology and consumerism	sign value, use value, media consolidation, consumption, technology fetish, identity, data warehousing, personalization
Technology and the body	reproductive technologies, genetic testing,

Technolog y and communiti es	cloning social media, digital divide, online gaming communitie s, community wi-fi
Technolog y and deviant social behavior	cyberstalkin g, identity theft, sexual predators, virtual rape
Technolog y and work	electronic monitoring, global workforce, deskilling / upskilling

Table 4: Learning Module

In addition to inclusiveness, the educator seeks to leverage the cultural diversity that exists in the classroom. Being socialized and living in the dominant culture often lessens our awareness that our beliefs and behaviors reflect a particular racial, ethnic, or gender affiliation. According to Wlodkowski and Ginsberg [15], white males have a more difficult time acknowledging the idea that Anglo-Americans and Western norms enjoy a privileged position in the United States' educational system. Culture is taught and understood implicitly, which is why it is difficult to explicitly describe who we are culturally. Consequently, we are more likely to experience uniqueness of our own culture when we are in the presence of those who are different from ourselves. Since, on average, 85% of the students enrolled in the course are white males, the challenge is to exploit the cultural diversity that exists within this group.

For example, students vary in their political affiliation. The social media learning activity exploits this difference by asking students to compare and contrast the use of You Tube, Facebook and other social media applications by 2008 presidential candidates. To help students understand how gender mediates social interactions, male students create female avatars and interact with others in a virtual world. They then report on their experiences. In the assignment on digital divide and internet indispensability, students keep journals on their internet use for a 24 hour period. They then reflect on the benefits of their use as well as the losses they may have suffered if they lacked internet access. After they complete this assignment, I ask students to collect statistics about ICT diffusion and other e-Readiness indicators for a developing country of their choice and compare these to comparable indicators for the US. Students are often astounded by the disparities that they uncover, and in some

cases express a deeper appreciation for the privileges that they derive from internet access. A few students are highly motivated by this exercise and extend this work for their semester long research project. These students conduct an experiment in which they give up their cell phones on for a weekend and reflect on this experience as a means of experiencing life as a "have not". Students consistently note extreme difficulty in coordinating schedules with friends, missing important phone calls from employers, the loss of social status that comes from cell phone ownership, and the loss of social connectedness that results from the inability to communicate with friends anytime / anywhere. One student recounted how he often provides technical support to his friends. During the weekend when he turned off his mobile phone, he missing a late night call from a frantic friend who experienced a computer problem while writing up a term paper that was due the following morning. As a result, the student was faced with the tough task of repairing his relationship with his friend.

There are a number of procedures used for classwork and assignments. Peer teaching is used in nearly every class. Students sign up to deliver individual presentations that correspond to the topics that will be discussed during the semester. In these presentations, students find current news articles, interesting websites, YouTube videos or other media related to the topic that they chose. Each presentation ends with a class discussion in response to two probing questions developed by the presenter. Through this exercise, each student acts as a teacher for the group. Additional procedures include group writing using wikis, journaling using blogs, small research projects such as examining privacy and identity using Facebook profiles, deconstructing advertisements for technology-related products, and using discussion forums to generate insights about characteristics of members of online gaming communities. Team-based research projects are the main way for students to demonstrate in-depth learning experience with humans and ICTs. Through these procedures, students are engaged in critical questioning, problem posing, and research that enable them to utilize their strengths, experiences, and values to deepen understanding of social, political, cultural and ethical issues that exist in the information environment.

34. IMPLICATIONS FOR EDUCATORS

The social justice approach has several implications for educators. First, educators have to address two sources of power asymmetry in the classroom – (1) the educator's institutional power, status and authority, and (2) the educator's social status characteristics such as race, gender, age and nationality. As an African-American woman teaching undergraduate courses to a predominantly white male student population, there are clearly many ways in which the educator and students are not alike. The challenge is to foster a classroom environment in which the educator's authority and expertise is clarified and made explicit because it bears on grading, evaluation, and classroom norms. Omolade [16] argues that "authority with, not authority over" is

complicated by social status differences between the students and the educator. Female educators of color must maintain democratic structures to foster authority with students. Once this mutual respect is established, students will accept greater responsibility for establishing an active and engaged learning environment.

Second, the instructor must create a safe environment that encourages discussion. Silence in the classroom can result from a number of factors including fear of being patronized, anger, anxiety, hostility, ignorance, or resistance to being forced to speak. For hooks [17] silence is not always about the failure to make an utterance. Sometimes, silence is the fear of being unable to make a speech that compels others to listen. Silence can be the fear of not being heard. Educators can use their classroom authority to create a safe classroom environment that compels students to value their own voice by establishing classroom norms to ensure respect and the amicable resolution of conflicts. Technologies such as message boards where students can post comments and pointers to relevant information provide another means by which students can participate in the course.

Finally, educators must enable students to understand that the power relations that exist in society are not the natural order of things. The privileges that we enjoy in life are the result of historically constructed forces which can be changed. As future information technology professionals, students can use the political influence, resources and human agency to create and implement ICTs that are socially just, environmentally benign, and enable positive social change.

35. CONCLUSION

Social justice provides a useful approach for infusing ethics and multiculturalism in undergraduate informatics education. A social justice pedagogy for the study of ICT and its social implications would necessarily include teaching that pays attention to personal safety and classroom norms, begins from students' worldview, and uses students experiences as a basis for course content and problem-posing. This student-centered approach is vastly different than the traditional lecture-discussion-testing mode of higher education that has existed for centuries, and requires a shift in the power relations between students and educators. Educators seeking to adopt a social justice approach may do so in incremental steps by experimenting with different assignments and classroom activities, and by fostering a classroom atmosphere that invites everyone to speak from their own voice of experience, take risks, and learn from their mistakes. Social justice education has the potential to prepare students who are (1) able to critically evaluate the intended and unintended consequences of ICT adoption and use, (2) sophisticated in their understanding of social institutions in shaping the meaning and intended uses of ICT, and (3) able to work with diverse others.

36. REFERENCES

- [54] Adams, M., Bell, L.A., and Giffin, P. 1997. Teaching for diversity and social justice: A sourcebook. Routledge, NY.
- [55] Friere, P. 1970. Pedagogy of the oppressed. Seabury, NY.
- [56] Hoesle, V. 1992. The third world as a philosophical problem. *Social Research*, 59, 227-263.
- [57] Paterson, B. 2007. We cannot eat data: The need for computer ethics to address the cultural and ecological impacts of computing. In *Information technology ethics: Cultural perspectives*, Hongladadrom, S. and Ess, C. (eds.), Idea Group Reference, Hershey, PA, 153-168.
- [58] Mbarika, V., Payton, F., Kvasny, L. and Amadi, A. 2007. IT education and workforce participation: A new era for women in Kenya? *The Information Society*, 23, 1, 1-18.
- [59] Hacker, K. and Mason, S. 2003. Ethical gaps in studies of the digital divide. *Ethics and Information Technology*, 5, 2, 99-115.
- [60] Kvasny, L. 2006. Let the sisters speak: Understanding information technology from the standpoint of the 'other'. *The DATA BASE for Advances in Information Systems*, 37, 4, 13-25
- [61] Tavani, H.T. 2004. Ethical reflection on the digital divide. *Journal of Information, Communication and Ethics in Society*, 1, 2, 99-108.
- [62] Ess, C. 2005. Lost in Translation? Intercultural dialogues on privacy and information ethics. *Ethics and Information Technology*, 7, 1, 1-6.
- [63] Westin, A. 2003. Social and political dimensions of privacy. *Journal of Social Issues*, 59, 2, 431-453.

- [64] Capurro, R. 2005 Privacy: An intercultural perspective. *Ethics and Information Technology*, 7, 1, 37-47.
- [65] Johnson, D. 1997. Is the global information infrastructure a democratic technology? *ACM Computers and Society*, 27. 20-26.
- [66] Winner, L. 1997. Cyberlibertarian myths and the prospect for community. *ACM Computers and Society* 27, 14-19.
- [67] Giroux, H. 1992. *Border Crossings: Cultural Workers and the Politics of Education*. Routledge and Kegan Paul, NY.
- [68] Wlodkowski, R. and Ginsberg, M. 1995. *Diversity and Motivation: Culturally Responsive Teaching*, Jossey-Bass Publishers, San Francisco, CA.
- [69] Omolade, B. 1987. A black feminist pedagogy. *Women's Studies Quarterly*, 15, 3 & 4, 32-39.
- [70] hooks, b. 1994. *Teaching to transgress: Education as the practice of freedom*. Routledge, NY.

FULL PAPER SUBMISSION 18:

Community Informatics

AUTHOR(S):

Williams, Kate
Durrance, Joan
Rosenbaum, Howard

ORGANIZATION(S):

University of Illinois, United States of America
University of Michigan, United States of America
Indiana University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

This proposal is for three papers which would form all or part of a panel presentation reporting current work in community informatics at three i-schools. It complements the wildcard proposal submitted on community technology/information/informatics, which we are each also part of. If there are 1-2 other paper proposals in this area we would be glad to be combined in one session with them.

As per the request for anonymity, the paper numbers correspond to the authors index in the webform.

All three of us will present.

1 Paper: TOP Data Archive: Repurposing data to advance the emerging field of community informatics

When a federal program which funded 600 community informatics projects over 11 years was shut down, researchers and staff worked to repurpose their files into a research dataset. Organizing the physical-plus-digital archive into a distributed collection proved to be effective teaching and research at two i-schools and one library school. A number of scholars across several disciplines are now using the dataset for their investigations. To our knowledge, there is no other shared community informatics dataset.

Analysis of the TOP data is deepening our understanding of several questions, among them how social capital and social networks interact with technology in local communities, and allowing us to inform policymakers interested in TOP 2.0. The presentation will report the methods used and the results thus far.

2 Paper: Identifying Community-Centered Outcomes

Nonprofits still struggle as they attempt to effectively identify specific outcomes of their services that accrue to individuals, their families, community groups, and the larger community. These skills are essential both to community-focused ICTs, libraries and other nonprofit staff and to students who seek to build their careers in these institutions. Faculty at iSchools that have embraced community informatics are now beginning to give students the knowledge and experience needed to help community organizations more effectively meet community needs and tell their story. This presentation will provide a contextual framework for identifying and presenting outcomes. More importantly it will provide examples of outcomes identified by graduate students who have conducted field-based outcome studies of a variety of community organizations. Outcome categories include increasing knowledge and skill levels, personal efficacy gains, fostering social capital, etc.

3 Paper: Theorizing community informatics: Memory practices in online communities

Based in the insight that digital archiving is a fundamental process in the sciences, Bowker (2005) has proposed the concept of memory practices as a way to understand a new regime of information and communication technologies for recording and storing memory. He argues that these practices are technical and social and include shared standards, meanings and protocols; they allow data and information to be stored and forgotten and then retrieved and remembered. In his view, they are routine components of work practices, imposing order on things, holding and shaping experience. This paper explores the utility of memory practices, seen here as technologies of remembering and forgetting, for understanding the infrastructures on which online communities depend.

FULL PAPER SUBMISSION 19:

“Just Me in a Closet with a Computer”: Student Perceptions of Undergraduate Information Technology Programs

AUTHOR(S):

Choi, Heekyung

ORGANIZATION(S):

University of Illinois at Urbana-Champaign, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

1. Introduction

Recently, a number of new information technology (IT) education programs have emerged in U.S. colleges and universities, in addition to the traditional computer-related education programs. The new IT programs are being identified with titles such as Information Science, Information Systems, Information Technology, or Informatics. These programs have been developed on the basis that there is a niche between IT and the contexts in which the technologies are situated, in terms of both further research and teaching. Many of these IT programs are providing undergraduate level IT education programs, either as a major or as a minor.

While the IT programs in colleges have been gaining popularity, the identity of the IT programs still remains fuzzy, and the nature and scope are still being debated intensively. One reason may be that the new area of study is still young, and the previous studies have not had enough time to sufficiently address multiple perspectives from various stakeholders of college-level IT education.

One of the most important stakeholder groups in the emerging IT education programs is college students. However, most previous studies have rarely examined what expectations and concerns college students have regarding IT education. Therefore, this study proposes to shift attention from the perspectives of program developers to the perspectives of college students.

The big research question that guides this study is,

What are college students' perceptions and expectations for an IT education program?

By investigating students' perceptions and expectations, this study aims to contribute to a better understanding of the IT education programs and eventually help educators answer their question regarding what the identity of the IT education programs should be.

2. Methodology

As the development of the college-level IT programs is a new phenomenon and students' perceptions on these programs have rarely been studied, there is no strong theoretical framework that can guide this study to answer the above question. This lack of a strong previous research framework requires some exploratory research that identifies the major issues first. This exploratory research was performed using data obtained through qualitative in-depth interviewing with college students.

Research setting: The research setting for this study is an undergraduate IT minor program at a large Midwestern university. While some IT programs were developed by the traditional computing disciplines, this IT minor program was originally developed by the Graduate School of Library and Information Science (GSLIS). Currently, the IT minor program is undergoing a significant transition, being expanded as a campus-wide minor rather than a GSLIS minor. It is closely collaborating with the Computer Science department on the same campus. The program will use a new name "Informatics minor" from the Fall 2008 semester, to indicate that this is a comprehensive campus-wide minor, which is not bound to a specific academic discipline. The curriculum is being restructured, and new classes are being added. The IT minor courses are open to non-minor students as well, and the introductory core courses fulfill the general education requirements of the campus.

Interviewees: College students who were enrolled in one or more courses in the IT minor program were recruited for the interview. The purposive sampling method has been adopted to obtain participants from various majors. Enrollment in an IT course was taken as evidence that indicates students' interest in the new IT education. Total 26 students from the following majors were interviewed; English, History, Advertising, Sociology, Accounting, Finance, Computer Science, Computer Engineering, Chemical Engineering, Aviation, etc.

Most interviews were performed during the Spring 2007 semester. Instructors of the courses were contacted first to obtain opportunities to recruit interview participants from their classes. The students in the classes were given a brief introduction of the study and asked to sign up for an interview session. Participation in the interview was completely voluntary. Each interview session lasted about an hour.

3. Findings

3.1. What does IT mean to students, as an area of study in college education?

- IT is about computers: IT broadly means any type of digital technology that processes information. But to students, IT in the context of college education primarily means "computers" - in most cases they mean personal computers, rather than other popular consumer IT device such as cell phones and music players.

- IT is about Web technologies: From the interviews, it is obvious that students' interest in IT is something different from their interest in computers themselves. Interview data indicate that students' interest in IT often has started with their experience with the Internet, especially with the Web technologies.
- IT is something that students will have to acquire knowledge about for their future, if they like it or not, or if they are good at it or not: Many students from various majors, from English to Computer Science, commonly perceive that "IT is not too exciting but something in which they will need to learn for their future".

3.2. Three major interests in IT education

From the interview data, three major fields of interest that college students have about IT education were identified.

- Designing/creating practical technical things on their own.
- Learning about theories about current affairs and social implications of IT.
- Interest in Human-Computer Interaction.

Interesting patterns were found among the major interests. The first two interests tend to be negatively correlated. Those who are more interested in practical hands-on stuff tend to be less interested in discussing social implications of IT, and vice versa. Also, some common underlying interests were found to all three interests. First, students are all interested in the practical aspect of college education. Second, all the three interests are focused around user-level issues; students take IT classes to make their websites, to make their own portfolio, to build their resume, to conduct their group project, or to learn how IT implementation affects end-users like themselves.

3.3. Why not the existing IT-related programs?

While many students have tried a CS class, they reported that the CS classes did not satisfy their needs. The reasons that the existing CS program is not satisfactory for them include;

- Some people are intimidated even without having direct experience with CS
- The strong negative stereotype of CS work, working long hours alone
- They feel that CS is very theoretical, for example, designing everything from scratch and producing outcomes that are not immediately usable.
- Students say that the class atmosphere is not very friendly to novices and less-experienced people.
- There is no consideration on the human side of computing.

Similar comments were made about the MIS program.

3.4. Perceived gap in college education

Students reported that they feel some gap in current college education in general.

- Most existing programs are very theory-based and not practical.
- Most programs focus on "history", rather than on current affairs and recent developments.
- There are rigid boundaries among existing disciplinary programs all across the campus. Many classes are open only to the students enrolled in a specific program.
- There is no class that is entirely devoted to discussions on IT.

3.5. Reasons for choosing the IT minor

While students choose their major based on what they like and what they are good at, they choose the IT minor based on its practical value for them.

- To supplement their major field, for example, philosophy plus IT
- To narrow down their expertise to an IT-related subfield, for example, to be a lawyer specialized in the cyberspace law
- To differentiate themselves from others with similar qualifications

3.6. Challenges for the IT minor program

Students' comments about the IT education program indicate that there are many challenges in building an undergraduate level IT education program.

- Most students' favorite way of learning IT is learning by themselves, by trial and error. If they prefer learning by themselves to learning with outside help, what would be the role of an IT education program?
- Unlike a major, the IT minor is competing with other valuable opportunities that students can take, including internships, study abroad opportunities, part-time jobs, campus honors' programs, and participating in student organizations.

- Balancing the technical and non-technical aspects in IT classes is definitely a challenge, as all students have different level of IT-related experiences.
- The negative stereotype about IT jobs, such as supporting technician's role, hard work, and low job security, affects students' willingness to join the IT minor.

4. Conclusion

This study reveals that many college students are looking at the void in college education, in terms of the recent developments of IT. This study also shows why the IT education program as a minor, rather than a major, makes sense to students. By taking the IT minor, students try to differentiate themselves within their own fields. The findings indicate that marketing of IT programs should be directed not only to students, but to various stakeholder groups whose perspectives have significant influence on students.

Finally, most interviewees were eager to provide their input to the IT program developers. This study is valuable in that it delivers students' perspectives on IT education to the IT educators' community.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

"Just Me in a Closet with a Computer":

Student Perceptions of Undergraduate Information Technology Programs

Heekyung Choi

University of Illinois at Urbana-Champaign

501 E. Daniel St.

Champaign, IL 61820

hchoi8@uiuc.edu

ABSTRACT

Many institutions of higher education in the US are offering an information technology (IT) education program at an undergraduate level. So far, not many studies about this young academic program have explored students' perspectives. Students' perspectives are important input for educators to implement an IT education program that properly addresses students' needs. This study has investigated college students' perceptions and expectations for formal IT education in college through in-depth qualitative interviews. Interviewees were college students who were enrolled in courses offered from an IT minor program. This study has revealed that many college students feel that there is a void in college education in dealing with issues emerging from recent developments of IT. The interviewees indicated that the existing computer science program did not satisfy their needs. Students perceived that IT is something that they would have to confront in their future, and sought IT education regardless of their fondness for using and learning IT. By taking the IT minor, students tried to supplement their knowledge in their major and be better prepared for the fields that they would pursue after

graduation.

Topics

Nature and scope of iSchools and iResearch

Keywords

IT education, IT learning, IT minor, computer science, programming, informatics, student perceptions

37. INTRODUCTION

Recently, a number of new information technology (IT) education programs have emerged in U.S. colleges and universities, in addition to the traditional computer-related education programs. The new IT programs are being identified with titles such as Information Science, Information Systems, Information Technology, or Informatics. These programs have been developed on the basis that there is a niche between IT and the contexts in which the technologies are situated, in terms of both further research and teaching. Many of these IT programs are providing IT education at an undergraduate level, either as a major or as a minor.

While the IT programs in colleges have been gaining popularity, the identity of the IT programs still remains fuzzy, and the nature and scope of IT education are still being debated intensively. One reason may be that the new area of study is still young, and the existing studies have not had enough time to sufficiently integrate multiple perspectives from various stakeholders of college-level IT education. One of the most important stakeholder groups in the emerging IT education programs is college students.

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse this work. The authors are solely responsible for their

However, most previous studies have rarely examined what expectations and concerns students have regarding IT education. Instead, they have discussed educators' and program administrators' concerns in developing and administering an IT program. Therefore, this study proposes to shift attention from the perspectives of program developers to the perspectives of college students to get a better understanding of college level IT education. The big research question that guides this study is formulated as follows.

What are college students' perceptions and expectations for an IT education program?

By investigating students' perspectives, this study eventually aims to help educators answer the questions regarding the identity of the college-level IT education programs.

38. RELATED STUDIES

College Students: The IT-savvy New Generation

In a large body of literature, younger generation people are typically described as more tech-savvy than older generation people. The digital divide literature has long been discussing the inequality that exists in people's access to IT. It points out that people's age and education level are two important factors that make a big difference in their access to IT (e.g., National Telecommunications and Information Administration [4]; Hargittai and Hinnant, [3]). College students are considered to be a group of people who have the most access to information technologies, as they belong to the younger generation that has grown up with IT and is receiving higher education that often goes together with intensive use of IT.

Spontaneous Learning

A significant improvement brought by the development of user-friendly information technologies is that those technologies have become easier to learn. Nielson [5], a usability consultant, points out that learnability, the ease of initial learning, is one of the top attributes in any good user-oriented system. One consequence of the enhanced learnability is that user' spontaneous learning is greatly encouraged. Many people acquire IT skills and knowledge spontaneously, through informal learning embedded in their daily activities, rather than through formal training or written manuals.

In contemplating how people learn to use computers, Twidale [9] pays special attention to users' informal help giving interactions that are frequently observed in workplaces. His idea of "over-the-shoulder learning" draws on education theories that students learn cognitive skills most effectively through informal, spontaneous learning opportunities embedded in natural classroom settings. According to these theories, interaction within peer relationships is essential when students learn complex cognitive skills (e.g., Vygotsky [10]; Rogoff, [6]). Other researchers, such as Splitter [8], also point out that a critical part of IT learning usually occurs in

users' social interaction with peers rather than through formal training.

The Place for Formal IT Education in College

As discussed above, college students are very active users of IT, and they acquire IT competency spontaneously in their daily lives rather than through any formal training. Previous studies suggest that the immediate value that learners attach to learning computers should be considered in thinking about IT education. Facer, Sutherland, Furlong and Furlong [2] point out that there is a fundamental gap between the reason that educators and parents place emphasis on computer education and the reason that children value computer expertise. They argue that while policy makers and parents tend to view computer skills in terms of the children's adaptability to the future world of work, the actual reason that young children value IT expertise is that obtaining the expertise is instrumental in shaping their identity among their peer groups.

College students have different values from those of young children who are the target of Facer et al.' [2] study. College students are situated in a transitioning period from education to "the real world". Therefore, college years are the time when career-related decisions are important to students. Developmental psychologist Smolak [7] argues that occupation-related values are a significant part of college students' identity formation. It implies that the reason that college students seek formal IT education may be closely related to their career-related issues. Also, many related studies have focused on computer skills, rather than on broad IT education which includes many conceptual parts of IT knowledge other than just computer skills.

If college students are generally savvy with information technologies, and they acquire IT competency more through informal paths, then what would be the place for formal IT education for them? What do they think they are not good at, and what causes them to consider formal IT education?

39. METHODOLOGY

As the development of the college-level IT programs is a new phenomenon and students' perceptions on these programs have rarely been studied, there is no strong theoretical framework that can guide this study to answer the research question. This lack of a strong research framework requires some exploratory research that identifies the major issues first. In this study, this exploratory research was performed through qualitative in-depth interviewing with college students.

Interview participants were college students who were enrolled in one or more courses in the IT minor program in a large Midwestern university. The purposive sampling method has been adopted to obtain participants from various majors. Enrollment in an IT class was taken as an indication of students' interest in IT education. A total of 26 students from the following majors were interviewed: English, History, Advertising, Sociology, Business Administration, Computer

Science, Computer Engineering, Chemical Engineering, etc. Interviewees included 15 male students and 11 female students. 3 sophomores, 8 juniors, and 15 senior students were interviewed. There were no freshman students, as the IT classes require sophomore standing. For confidentiality, pseudonyms were used to identify the interview participants throughout the study.

Interviews were performed during the Spring 2007 and Fall 2007 semesters. Students in the IT classes were given a brief introduction to the study and asked to sign up for an interview session. Each interview session lasted about an hour. The interviews were conducted in a semi-structured format. The semi-structured format allowed the interviewer to adjust interview questions flexibly to get in-depth understanding of each interviewee's experiences with IT and meanings that they attach to IT education.

40. FINDINGS

Major Interests in IT Education

From the interview data, three major areas of interest that college students have about IT education were identified.

Learning practical hands-on IT skills

The interview data revealed that one big interest that college students have regarding IT education is learning practical hands-on IT skills. IT broadly means any type of digital technology that processes information. But to students, IT in the context of college education primarily means computers, rather than other popular consumer IT devices such as cell phones and music players. It seems that their interest in personal computers is related to creating practical digital artifacts on their own. These digital artifacts include their personal web pages, portfolios for jobs, and tools to support the groups that they belong to. Emily, a psychology student, expressed how excited she was when she learned some practical hands-on IT skills from an IT class.

Emily: The things that we created in the class, it was something that actually worked and could actually be used. We created things that I feel that we accomplished something, which is kind of cool.

Creating something on computers usually involves programming. However, students clearly indicated that what they are interested in is not the programming per se, which CS people would do. Their interest in programming is limited to some types of Web programming that are needed to make their own stuff.

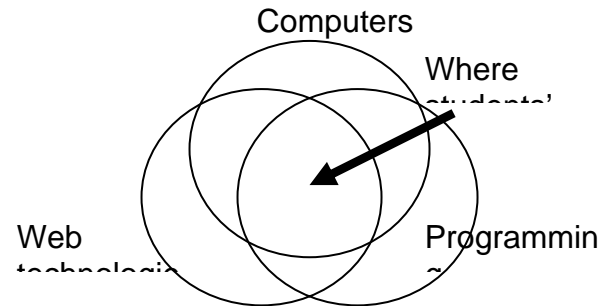


Figure 1. Interests in learning practical hands-on IT skills

Learning about social implications of IT

Students also showed interest in learning about implications of IT use on people and society. They are interested in learning how people use IT, and what implications IT use has for society, but not necessarily interested in using IT products proficiently or learning hands-on IT skills. Two students, from English and Business respectively, described their interest in learning the social implications of IT implementation.

Jenna: My expectations were not so much relating IT classes to technical stuff...but more relating it to culture... and current events more like, how the computers have evolved.... All I really thought about it was, I didn't really think it would be too technical.

Rachel: I've always liked learning about technologies and liked how people use it and how it affects people. I don't necessarily like using technology itself. I took a class last semester.... It was a communications class, and it was a lot about TV, Internet, and different types of media and how they affect society and different groups and different stereotypes.. I thought it was interesting. So I thought that the IT minor is fairly similar to that.

These comments indicate that even students who are not very active in using IT products feel the need for some kind of IT education. Their focus on the impact of IT is clearly different from the focus of the established approaches of CS or other engineering fields.

Interest in Human-Computer Interaction

Human-computer interaction (HCI) is another area that many students showed interest in regarding IT education. Chris, a philosophy major, focuses on computers less as technologies, but more as people's tools for communicating.

Chris: I am more interested in computing as interaction, less than computing as a goal and end itself. If people are interested in studying how people interact with computers, how computers shape their lives, the IT program is an ideal place to investigate some of that knowledge.

Students perceive HCI as an important issue to pursue in higher education, because it is related to many current changes in society.

Jack: That class [Design Information Interfaces] was very helpful, because I think a lot of things are either on the web right now or going in that direction. I think usability is really important because I've been on so many websites, not being able to find a button or not knowing where to click, and that's one of the big issues, it just does not allow people design for usability.

Some students also view HCI as a gateway through which non-technical students can get into the study of IT.

Jenna: Because I am not really very experienced with computers, I asked the IT advisor what classes would be less technical, and more underlines of computers and culture, human interaction with computers...

It seems that HCI is a major area of IT education. By HCI, students mean not a narrowly defined sub-area of IT, but rather people's interaction with computers in a broad sense.

Patterns found among the major interests
Interesting patterns were found among the major interests. The first two interests, interest in learning hands-on IT skills and interest in learning social implications, tended to be negatively correlated. Those who were more interested in acquiring practical hands-on computing stuff tended to be less interested in discussing social implications of IT. Those who showed more interest in learning about social implications of IT tended to be less interested in learning technical stuff. Also, some common underlying interests were found in all three interests. First, most of the students showed interested in the practical aspects of IT education. Second, all the three interests were focused around user-level issues; students take IT classes to learn to make their websites, to make their own portfolio, to conduct their group project, or to learn how IT implementation affects end-users like themselves.

"I am not good at computers": Conflicting Attitude towards their IT competency

Many students had conflicting attitudes regarding their IT competency. While they said they use IT all the time for their school work and socializing purposes, many of them rated their overall competency in using IT not very high, with a comment like "I am not good at computers". It seems that their intensive use of computers for school work and social purposes does not count as IT competency for them. A lot of times the feeling of incompetence came from their inability to do programming.

Jack: Technology, I do utilize it a lot. I think I'm pretty familiar with a lot of the IT technology in terms of the social use... utilizing different technologies.... The one thing that I don't feel comfortable with is the coding itself.

It is noteworthy that students did not describe their IT competency as a standalone technical skill without a context. Most times, students described their IT competency in relation to some area expertise. For example, a journalism student Abby indicated that she is not competent to sit down and write a program by herself, but that her IT skills would be good enough to perform her work in the journalism area.

Abby: For the future, I'll need to know more about online journalism. But could I sit down and make a web page, no, I think I would need more interaction. I think I have enough base knowledge to accomplish anything that I would want to accomplish.. if I want to learn more, I could figure it out on my own without struggling too much.

It seems that students feel that their IT competency is good enough if their IT skills and knowledge are functional in performing what they need to do.

Perceptions about the Existing CS Program

The interview data revealed that many of the interviewees have some sort of interest in learning issues in computing. As a result, many students have taken or considered taking a CS class. Most students reported that the CS classes they tried did not satisfy their needs.

First, students generally perceived that computer science is very hard. CS classes are usually considered to be hard, because they require a lot of math background. Some students pointed out that CS classes are daunting even for people who already have some level of technical competency outside of class.

Chris: I joined the university with very sporadic knowledge of computing. I had fairly good knowledge of coding and software systems.... However, computer programming courses in computer science department are notorious, being exceedingly difficult even for people who already know the basics of the programming.

Students also commented that the class atmosphere of CS classes is not very friendly to novices and less-experienced people.

Second, the negative stereotype of CS work influences students' perceptions of CS education. Many students believed that CS work typically requires long hours of working alone with tedious tasks, and pointed to this as a very unattractive aspect of learning programming.

Dave: Programming is a good skill to have. It's not something like for sure going for career, no way. I thought it's too dull for me personally, because it's solitary. I feel like I have more social skills, I would rather talk and interact with people on group work, than have just me in a closet with a computer.

Third, they feel that what is being taught in CS is very theoretical rather than practical. They pointed out that CS classes typically require designing everything from scratch and producing outcomes that are not immediately usable. This is not practical for most students who are not CS majors, because the CS way of programming is not only hard for them, but also takes a lot longer time than necessary to get what they want.

Kelly: My friends have had taken them, they come home talking about their really arbitrary assignments, things that are not gonna help them. I didn't want to deal with those assignments. I want to learn how to really get end up and do something that I can actually physically use, create a product. I know that you need to learn that in order to get to that. But that's not what I need....

Lastly, some students expressed their concern that CS does not consider the impacts of computing on human life.

Chris: One of the things that I found which is really good about the IT program was that it really focuses more on the interaction between people and computers, instead of just the very low-level programming aspect, which is something that I really missed in the computer science department. There wasn't really anything that computer science can offer me that I was really looking for.

Students' comments about the CS program indicated that the focus of existing computing education does not go very well with non-CS students' needs for computing education.

Perceived Gap in College Education

Students reported that they feel some gap in current college education in terms of education for IT. First, students pointed out that most existing academic programs are very theory-based and not practical. Nathan is a computer engineering major. He described his experience that all the theoretical knowledge that he had obtained from his study was not very useful in performing his work assignment.

Nathan: I do like a lot of practical things, and am not real big into studying theories. I don't mind studying a theory to get an understanding of how software program works, like the course I had last semester, we used a computer design tool to help us lay-out a processor. So we actually made our own little processors, and one of things you had to connect a bunch of different wires between all these different little pins. If you do it manually, you will have a really big problem, you can use a choice of different algorithms to automatically do it for you. So I learned about different algorithms in class and used them in a practical way. That's about the technology that I'd like... where I can say "All right, I can use this." "Here's an algorithm, you will use it maybe!" I am like, "okay..." Because I am not really big into theory, and just kind of

good at research on that algorithm, it's not all that useful...

Second, students commented that most academic programs focus on "history", and are slow at embracing issues of current affairs and recent developments. This complaint appears over a wide range of academic disciplines.

Dave, a political science major, mentions that political science deals more with historical stuff than the changes that he is experiencing right now.

Dave: Courses in my major deal a whole lot of government stuff, seem more of just history, looking back at the history of government. While in this [IT] course, I feel like what we are studying is going right now. I enjoy that, because it's something that I feel like I am in actually a part of, because I am participating on the Internet which mostly these classes talk about, and how information is shared through different chat. I feel like it's not just studying the past... it's studying now.

Another criticism was that there are rigid boundaries among existing disciplinary programs across campus. Many classes are open only to the students enrolled in a specific program. Many social science classes deal with IT issues at an upper-undergraduate level. These upper-level courses usually require several prerequisite courses within a specific discipline. There is no social science class that is entirely devoted to discussions about IT.

Zachary, a CS major, is taking an IT course that fulfills his general education requirements for social sciences. He is very much into his study in CS, and is also interested in different kinds of discussion about computers and other technologies. As he did not want to go too far away from technical issues when choosing social science courses for his general education requirements, he decided to pick two IT courses that satisfy the general education requirements. He found that it is difficult to find such courses from other social science disciplines, because other disciplines usually offer technology-related courses only at a higher level for junior and senior students in their fields.

Zachary: [If I am looking for a social science course that deals with IT,] that'll be five courses to get to anything that really about computers instead of two [introductory IT] courses.

As IT is increasingly being adopted in many settings in everyday life, students do have some sort of sense that there might be some education program on IT on campus, even without explicit knowledge. This is expressed in the conversation with Brad, an economics major who is not very fond of working with IT.

Heekyung: Did you know that there is an IT minor on campus?

Brad: I guess I would've.... I don't know explicitly, but I figured that there would be...IT isn't something that I'm extremely interested in, so I wouldn't have considered it, I wouldn't look into finding whether there was, but yeah, I guess I didn't. Not explicitly, but I would have if it was something I was interested in, I would have figured that you can get an IT minor. I had this feeling that there is something like an IT minor on campus.... I figured it would be there, but I didn't actually know.

It seems that their perceived needs for IT education have to do with their perceived gap in current college education.

Reasons for Choosing the IT minor

While students choose their major based on what they like and what they are good at, the interview data showed that students choose the IT minor because of its practical value that IT education bears to them.

A few students reported that they chose IT courses because they generally like to work with IT. However, for many other students, IT is something that they will have to acquire knowledge about for their future, whether they like it or not, or whether they are good at dealing with it or not. Many students from various majors, from English to Computer Science, commonly perceive that IT is not very exciting, but something that they will need to learn eventually for their future life beyond education.

Kate, a humanities student, has been hesitant to embrace new technical things since she was young. However, she feels that it is something that she would need to confront in developing her future career anyway. Kate repeated that learning IT is very hard for her, indicating that she is taking the IT minor for its instrumental value, not because of her fondness of what is being taught in IT education.

Kate: Technical things are hard, but worth learning. Especially I am coming into the graduate school of library and information science next year, so I know that I am gonna need to know things about Internet and technology. It's just hard. It's hard for me.

Zachary, a CS major, also talks about the need to learn about IT regardless of his willingness to stick to a CS job. In his mind, a CS job is a job that focuses on programming, as opposed to an IT job, which is more about providing technical support to other people.

Zachary: I know I am most likely ending up with being a coder somewhere. I will have to use code repositories, knowledge bases... a lot of tech support things. They put that in a repository.

Students said that they chose the IT minor because they believed the study of IT supplements their major field well. Abby, a journalism major, explained her need to learn IT to be up front in her area of study.

Abby: I feel like it [the IT minor] complements my major, journalism, as where the technology of journalism is going, and what new mediums we will be using to portray the news. I do think they are hard, and I haven't understood everything they said, but I think overall I have a much better grasp of technology than I did

Another student mentioned that he is taking the IT minor to narrow down his expertise in writing to an IT-related subfield, for example, to become a lawyer specialized in cyberspace law.

Larry: I am going to law school, and I am planning on doing something very related to IT, or technology in general, maybe something like cyber law....

Another important reason is to differentiate themselves from others with similar qualifications. Kelly is a humanities student who is planning to continue her study in library and information science after college. She thinks that there are so many librarians-to-be who have similar qualifications. She considers IT competency as something that would make her unique by showing her well-roundedness.

Kelly: I feel like it'll at least give me a little bit more a well-rounded aspect just than just doing humanities. Having any sort of IT competency is gonna put you so much higher in the rankings as far as finding positions than anyone else.

Most students' comments indicated that some practical value of receiving IT education, rather than an inherent interest in IT, is a critical factor in their decision to take the IT minor.

Factors that Influence Students' Choice of IT minor
Students' comments about the IT education program indicate that there would be numerous challenges in building and administering a successful undergraduate IT education program.

First of all, the IT minor is optional to students, unlike a major. Students mentioned that they have to weigh carefully the value of doing the IT minor with other valuable opportunities that they could possibly take, such as internships, study abroad opportunities, part-time jobs, campus honors' programs, and participating in student organizations.

Another important factor in doing the IT minor is how flexible the course requirements within their major are. While some social science majors do encourage students to have a minor with an additional requirement like "20 hours of supporting coursework", other majors leave no room for students to seek another interest. Nathan, who is a senior in computer engineering, cautioned that a student in his major has to be very careful in choosing to do something outside of his major.

Nathan: In engineering, if you're going to take courses that aren't technical, a lot of times you need to make sure if those are counting for something.

For him, if he takes something outside his major without confirming that the outside course counts towards something for his degree, it would cause trouble in finishing his degree on time.

Some students expressed their worries that other important people like hiring managers would not know what the IT minor is. Rachel, a business student, made an interesting comment, based on her interviewing experience for an internship.

Rachel: [In my interviews,] I asked some of the recruiters, what would they think of the IT minor. I think a lot of recruiters would misunderstand, they think IT as computer science, so they think oh, you're an accounting major, we need you to do accounting, we don't need you to fix computers. That's their idea. I don't think they would understand what it is.

"Fixing a computer" may be a popular stereotype about IT work. Rachel's comment signals that raising other stakeholders' awareness about an IT education program would be an important factor for students to have a positive impression about an IT program. If other people who have significant influence on students' future do not understand what an IT program is, it would make students hesitant about choosing an IT program.

41. CONCLUSION

In this study, students from different majors showed diverse interests in IT education. However, the existing computer science program does not seem to respond to those interests. Students also pointed out that higher education in general has a gap in addressing the issues emerging from increasing implementation of IT. In addition, most students indicated that they are anticipating some practical advantages by enrolling in an IT education program. The findings also signal that marketing of an IT program should be directed not only to students, but to various stakeholder groups whose perspectives have significant influence on college students.

Students' comments also implied that IT education is facing many challenges. It seems that many of the challenges for an IT program come from its interdisciplinary nature. IT programs are usually built as an interdisciplinary endeavor, as they usually start with the recognition that there is a niche between technologies and their social contexts. To fill the niche, the IT programs emphasize gathering multiple perspectives for viewing computing, from sociology, from communications, and from computer science, etc. To sustain an interdisciplinary program, gathering and organizing supports from many disciplines is crucial. Berghel and Sallach [1] have already observed that many IT programs are being organized at a university level rather than at a departmental level.

The scope of this study is limited to an IT minor program in a specific university. Therefore, the findings may not hold for students in other universities or in other IT programs that are offered as a major program. This case study was conducted with the intention of getting an in-depth understanding of students' perspectives. There are not many previous studies about students' perceptions about IT education. At this stage, the benefits of identifying students' various issues involved in receiving IT education outweigh the value of getting generalizable findings across a large number of college students.

The findings strongly suggest that students expect an IT education program that has a different focus than that of the other existing computing programs. "Just me in a closet with a computer" may be a common portrait of programming/computer science that students have. Wanting something different from that is a major reason that students seek IT education. Students from various majors want some computing education that is more interesting, more practical, and more meaningful than CS. This study is valuable in that it delivers this anticipation to the IT educators' community.

42. REFERENCES

- [71] Berghel, H., and Sallach, D. L. 2004. A paradigm shift in computing and IT education. *Communications of the ACM*, 47(6). 83-88.
- [72] Facer, K., Sutherland, R. J., Furlong, R., and Furlong, J. 2001. What's the point of using computers? The development of young people's expertise in the home. *New Media and Society*, 3(2). 199-219.
- [73] Hargittai, E., and Hinnant, A. 2005. Toward a social framework for information seeking. In Spink, A. & Cole, C. (eds). *New directions in Human Information Behavior*. New York: Springer.
- [74] National Telecommunications and Information Administration. 1999. Falling through the Net. Available at <http://www.ntia.doc.gov/ntiahome/fttn99/contents.html>
- [75] Nielsen, J. 1994. *Usability Engineering*. Boston: AP Professional.
- [76] Rogoff, B. 1990. *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.
- [77] Smolak, L. 1993. *Adult development*. Englewood Cliffs, N.J.: Prentice Hall.
- [78] Splitter, V. K. 2005. Learning to use IT in the workplace: Mechanisms and masters. *Journal of Organizational and End User Computing*, 17(2). 1-25.
- [79] Twidale, M. B. 2005. Over the shoulder learning: supporting brief informal learning. *Computer Supported Cooperative Work*, 14(6). 505-547.

- [80] Vygotsky, L. S. 1986. Thought and Language.
Cambridge, MA: MIT Press

FULL PAPER SUBMISSION 20:

The Social (Tagging) Act of Reading

AUTHOR(S):

Malone, Cheryl Knott (1)

ORGANIZATION(S):

University of Arizona, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

On Flickr, a Web-based service for managing and sharing digital images, reading and tagging come together in ways that reveal how people -- as individuals organizing their own photographs as well as making them available for discovery by others -- think about reading when they attempt to depict it in images and in words. In this pilot study of the tags assigned to a sample of 100 photographs tagged with the terms book and reading, only 80 of the 561 total tags were used more than once. The broad range of unique tags suggests an idiosyncratic approach to labeling that makes it difficult to draw conclusions about images of reading. Consequently, additional searches were conducted on Flickr and the results are reported and interpreted. The paper concludes with a discussion of proposed follow-up research.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

The Social (Tagging) Act of Reading

Cheryl Knott Malone

UA SIRLS

1515 East First Street

Tucson, AZ 85719

1-520-621-3565

ckmalone@u.arizona.edu

ABSTRACT

On Flickr, a Web-based service for managing and sharing digital images, reading and tagging come together in ways that reveal how people -- as individuals organizing their own photographs as well as making them available for discovery by others -- think about reading when they attempt to depict it in images and in words. In this pilot study of the tags assigned to a sample of 100 photographs tagged with the terms book and reading, only 80 of the 561 total tags were used more than once. The broad range of unique tags suggests an idiosyncratic approach to labeling that makes it difficult to draw conclusions about images of reading. Consequently, additional searches were conducted on Flickr and the results are reported and interpreted. The paper concludes with a discussion of proposed follow-up research.

Topics

Cultural information systems
Information organization

Keywords

reading, books, social tagging, Flickr, photographs

43. INTRODUCTION

Mitchell Stephens in *The Rise of the Image, the Fall of the Word* describes three places -- a physician's waiting room, a child's bedroom, and a passenger jet -- and asks "What's missing from these pictures?" His answer: "Reading is now missing from countless scenes it once dominated: at kitchen tables, on buses and trains, in beds at night, on couches, even in some classrooms."

, p. 8] Reading of literature in particular is less present in American lives, according to a National Endowment for the Arts survey, which in 2004 identified "an overall decline of 10 percentage points in literary readers from 1982 to 2002, representing a loss of 20 million potential readers." [6] At the same time, scholars of print culture have produced quantities of studies

of reading, both current and historical. The work on current reading practices often uses ethnographic methods to observe and interview readers while the historical studies draw on the customary sources of historians, archival and published texts, to reconstruct readers and their experiences. Their work documents the political, social, individual, and cultural functions, uses, and meanings of reading in diverse circumstances and locales.

Reading would at first glance seem to have nothing to do with collaborative or distributed tagging of information resources, especially images. Both are of interest to researchers and teachers of library and information science (LIS), albeit to different and probably not overlapping communities of interest, despite the fact that neither reading nor tagging is considered core to LIS paradigms or to the LIS curriculum [4], [9]. Print culture theorists understand reading as both a solitary and a social act. Similarly, social tagging of information resources has both individual and social aspects. One place where reading and tagging come together is Flickr, a Web-based service for managing and sharing digital images. Images tagged "reading" constitute a new body of documentation for print culture scholars, who typically have relied on the word to study the word, but they also offer a body of data for LIS researchers interested in understanding the phenomenon of collaborative indexing and what it can tell us about amateur or what Beghtol [1] has called "naïve" information resource description and organization. Corinne Jorgensen [4] has pointed out that there are now three ways to provide intellectual access to images: human-assigned controlled vocabulary terms, computer-generated indexing, and social (or distributed or collaborative) tagging. She suggests that social tagging may help reduce the "semantic gap" between human indexing with its interest in contextual meaning and automated indexing with its focus on describing the content and characteristics of the object represented in an image.

On Flickr, reading and tagging come together in ways that reveal how people -- as individuals organizing their own photographs as well as making them available for discovery by others -- think about reading when they attempt to depict it in images and in words. How people conceptualize and represent reading in the early 21st century matters to students of print culture interested in the continuing and changing activities and behaviors associated with literacy and to students of information interested in the

continuing and changing activities and behaviors associated with resource description for organization and retrieval. Consequently, this paper marks the beginning of a research project designed to uncover and analyze modern-day graphic and textual representations of reading in everyday life. The preliminary work reported here focuses on identifying what is depicted images labeled with the tags books and reading. This work will serve as a foundation for further research into the relationship between reading and tagging as social acts but also as an intellectual interaction characteristic of life in the Information Age.

44. METHODS

Approximately 55 million photographs are on Flickr. [3] The number fluctuates as new photos are uploaded and as existing ones are removed. A simple keyword search of the full text of Flickr for reading yields more than 700,000 results (after deleting those where the keywords pennsylvania, penn, and pa also occur in an attempt to eliminate photographs of the city of Reading). When the search is limited to tags only, the result is almost 109,000. This pilot study used two methods to reduce the number of results for the purposes of this report. The search query included the words reading and book limited to tags only, bringing the total results to 11,213. A random sample of 100 of these results was selected for the study. The 100 photographs depicted a range of compositions including individuals and groups handling books, still-life compositions involving books with other objects such as coffee mugs and eyeglasses, shots of individual books both open and closed, and stacks or piles or shelves of books. The researcher viewed each of the 100 images and their tags and recorded and tabulated the tags.

To address Mitchell Stephens's assertion about where reading occurs, additional searches were conducted using keywords such as train, bus, and airplane combined with book and limited to tags only. The point is not to prove that books still have a place in the Information Age as evidenced by their robust representation on a photo-sharing site, but to explore what constitutes "reading" to the collective mind of Flickr image taggers and to elicit questions for further investigation.

45. FINDINGS AND ANALYSIS

Although this study is limited to images tagged with both reading and book, many other keywords related to the topic are assigned as tags on Flickr. Table 1 lists some of them and the number of records with those tags (as of 01/24/08), ranked by most to least common. The singular form book is the most commonly used tag, but it is not clear that every image tagged with the term is using that keyword to mean the printed codex. What is clear is that some photographs tagged with the term book do not depict a book at all but merely evoke the idea of a book, such as the image of a young man looking down [2] or of a solitary bookmark [5]. Not surprisingly, more specific tags, such as novel and hardcover are used less frequently than broader tags.

Table 5. Reading-related keywords as Flickr tags

Keyword	Occurrences
Book	199,416
Books	137,162
Reading	108,553
Read	13,755
Novel	5,842
Reader	3,806
Paperback	2,999
Novels	1,713
Readers	1,597
Hardcover	814
Mybook	747
Reads	654
Mybooks	644
Hardback	479

Softcover	406
Chapbook	233
Chapbooks	22

Although the 100 photos included in this study are not necessarily tagged with the keywords shown in Table 1, they nevertheless are useful because they were tagged with two of the most popular ones: book and reading. The number of unique tags beyond book and reading assigned to these 100 photographs totaled 559. Of the 561 total tags assigned to the 100 photographs, only 80 were used more than once, as shown in Table 2.

The terms have not been grouped into categories, but some ways to do that are obvious. For example, it would make sense to combine the terms me, self, and self-portrait (which is already a combination of self-portrait and selfportrait) into a single group

Table 2. Additional tags for photos tagged book and reading

Tag	Occurrences
Read	9
Books	7
Me	7
self-portrait	6
Coffee	5
Girl	5
Library	5
People	5
Art	4
Café	4
Christmas	4
Glasses	4
portrait	4
Woman	4
2005	3
American tour	3
Author	3
baby	3
beautiful	3
Bed	3
Blue	3
bookstore	3
Bouchercon 2007	3
Country	3

crime writer	3
France	3
Glacier	3
Home	3
literature	3
Michael Jecks	3
novelist	3
Paris	3
Red	3
San Francisco	3
School	3
Signing	3
View	3
Writing	3
365 days	2
365days	2
Architecture	2
Beach	2
Black	2
Bookmark	2
Bw	2
bwPingu	2
California	2
Cat	2
Days	2
Display	2
Dog	2
Doll	2
Eslite	2
Eslitebooks	2
Explore	2
Family	2
Felt	2
Holiday	2
Illustration	2
India	2
lady	2
Lesson	2
Librarian	2
Libro	2
Livro	2

Man	2
Mom	2
Nikon	2
Peets	2
Reader	2
Self	2
Shoes	2
Sitting	2
Statue	2
Story	2
Street	2
Student	2
Studying	2
Sunset	2
Taipei	2

totaling 15 occurrences in the 561 tags used in the 100 photos. But the tag me, tied for second place with 7 occurrences, is interesting in and of itself.

It's commonplace for library and information science (LIS) practitioners and researchers to assert that tags are akin to subject headings and indexing terms, but the tag me arose out of social tagging and not out of formal, expert controlled-vocabulary construction. For those of us with an LIS background, the tag me seems superfluous at best and downright nonsensical given its ambiguity and its seeming lack of utility for information retrieval. And, yet, it has its uses, however unexpected. One Flickr user has written: "The 'me' tag is easily my favourite tag on Flickr, it shows the wide variety of different folks who post to Flickr, a wide variety of ages and genders, appearances and attitudes." [8] Tagging a photo me may serve as the photographer's way to take credit for their work without having to reveal their real name or other personal information.

Table 2 does not indicate what's most striking about the ways in which images related to reading are tagged because it does not list the 479 tags assigned to only one of the 100 photos in the sample set. These unique tags range alphabetically from A train to zig. Many of them are the first names of the individuals captured in the photographs. Others are the titles and authors of the books depicted. As with the tags in Table 2, such as library and studying, some of these single-use terms seem highly correlated to the reading of books. For example, the terms include literary festival and author event. Other terms, such as the active sailing and the passive sleeping, seem incompatible with the act of reading.

The tags for the sample of 100 images do not quite address Mitchell Stephens's assertion that it is rare to find people reading in public or private spaces. The sample set is small relative to the total number of images on Flickr. And people do not tag in a systematic manner requiring them to make sure every object shown in a photograph has a corresponding tag. But a search in all of Flickr combining the tag book with other tags representing a variety of settings did turn up some results worth considering, as seen in Table 3.

Beach emerged as the most frequent setting of the setting-related tags searched. It may be that people take books to the beach in greater numbers than to Starbucks. Perhaps they take newspapers to Starbucks instead. It could also be that people take books and digital cameras on the relatively rare and relaxed days when they visit the beach more often than they take both to Starbucks when they stop in on their way to work. In other words, each row in the table can only suggest some of the contexts in which books can be found. The table as a whole, however, provides us with a sense that some books coincide with some people's eating, drinking, loafing, and traveling.

And with their shooting and sharing of photographs.

Table 3. Photos with book tag and a contextual tag

Tags	Occurrences
beach	1,436
table	1,096
coffee	992
bed	929
park	788
train	594
kitchen	502
café	495
lunch	494
hotel	452
couch	439
bedroom	357
office	338
Starbucks	201
dinner	200
bus	176
pool	173
breakfast	170
metro	160
subway	159
airplane	158
waiting	117
coffeeshop	66

46. FURTHER RESEARCH

This preliminary study cannot offer definitive conclusions, but it can serve as the foundation for further research, especially because it can suggest additional questions to ask and methods to use.

Here are some research questions that have occurred to me and that I plan to pursue as a result of this early work.

- What demographics are represented in photos depicting people and books? What proportion show different age groups, genders, ethnicities, and nationalities? What are the methodological issues involved in trying to interpret demographics from images when the accompanying tags do not provide any demographic indications?
- Which books are shown in photographs tagged with terms related to books and reading? Which authors are represented, which specific book titles, and which categories of books are shown and in what proportion? What impact does a media event such as the release of a Harry Potter book have on the number and type of photos published?
- How do the titles and captions and comments supplement or subvert the tags assigned to photos? How does the photostream and/or themed set in which the photo appears alter the viewer's understanding of the photo?
- To what extent are photos of reading staged versus candid? To what extent is the book a prop? What is the function of the book as prop—for creating a particular self-image or for product (or author or bookstore or library) promotion or to evoke the centuries-old symbol of literacy or revolution or erudition?

The last set of questions will require surveying and/or interviewing photographers. The remaining questions can be answered largely through the use of good data gathering and analytic methods.

Flickr is a popular and successful social networking site and its features serve its users well. But the researcher needs different kinds of features and functions. The large number of unique tags assigned to photos, the lack of a highly developed search engine with full Boolean operations and truncation devices, and the absence of a controlled vocabulary of indexing terms hamper the researcher's ability to retrieve a reliable and accurate set of results depicting reading. Consequently, the data-gathering and analysis methods used in subsequent work on this topic should include researcher-generated indexing terms for each image being studied. This approach would introduce much-needed consistency; for example, a photo of someone reading in the aisles at Borders and another of a reader sitting in a chair at Barnes & Noble would both be assigned a controlled-vocabulary term such as "bookstore." This approach would also allow not only for greater consistency but also for comprehensiveness, at least for the researcher's interests. For instance, a photo of a person sitting on the grass in a city park would probably never be tagged "public" but the researcher interested in making a distinction between reading in public and reading in private would find such a tag useful. Finally, the researcher's work should be grounded not only in the literature of social tagging but in the theory of the social uses of photography and in work on the intersection of print and digital culture.

47. CONCLUSION

The practices of sharing digital images and providing collaborative tagging to describe images have created an intriguing body of work that can be mined for information related to many topics, not least of which is the topic of books and reading in the era of digital information. But neither images nor their labels offer unambiguous evidence for the persistence of the printed codex, its relative presence in people's lives, or the uses to which it is put. A preliminary look at a selection of digital photos tagged book and reading suggests that, with carefully crafted research methods, both image and word can be studied together to help us understand individual and social uses of reading, or perhaps more accurately, individual and social uses of representations of reading.

48. REFERENCES

- [1] Beghtol, C. 2003. Classification for information retrieval and classification for knowledge discovery: Relationships between "professional" and "naïve" classifications. *Knowl. Organ.* 30, 64-73.
- [2] Finraz. 2007. DSC_8547.jpg. DOI= <http://www.Flickr.com/photos/finraz/1815754754>
- [3] Flickr. Creative Commons. DOI=(<http://www.Flickr.com/creativecommons/>)
- [4] Jørgensen, C. 2007. Image access, the semantic gap, and social tagging as a paradigm shift. In Lussky, Joan, Ed., *Proceedings 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*, Milwaukee, WI. DOI=<http://dlist.sir.arizona.edu/2064/>.
- [5] MoonlitBeadery. 2008. Cobalt & Cranberry Baking Bookmark. DOI=<http://www.flickr.com/photos/moonlitbeadery/2209086765/>.
- [6] National Endowment for the Arts. 8 July 2004. Literary reading in dramatic decline, according to National Endowment for the Arts Survey. DOI=<http://www.nea.gov/news/news04/ReadingAtRisk.html>.
- [7] Stephens, M. (1998). *The Rise of the Image, the Fall of the Word*. Oxford University Press.
- [8] striatic. 2004. Tagography ~ case studies. FlickrCentral/Discuss. DOI= <http://www.Flickr.com/groups/central/discuss/2730/>.
- [9] Wiegand, W.A. 1997. Out of sight, out of mind: Why don't we have any schools of library and reading studies? *J. Educ. Libr. Inf. Sci.* 38, 314-326.

FULL PAPER SUBMISSION 21:

Patterns in Collaborating and Communicating Behaviors: A Research Agenda for Library and Information Science

AUTHOR(S):

Nilan, Michael; D'Eredita, Michael

Organization(s): School of Information Studies, Syracuse University, United States of America

ORGANIZATION(S):

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Patterns in Collaborating and Communicating Behaviors: A Research Agenda for Library and Information Science

Abstract

Since the introduction of the World Wide Web (April, 1995), we have seen dramatic changes in technology and its artifacts that we call information. These developments allowed for a plethora of collaborative environments and new means for communicating to emerge. These are a direct result of specific and ubiquitous human behaviors manifesting themselves in seemingly distinctive ways. This has increased the salience of a set of formidable and—we argue—distinguishing phenomena our field is now addressing. This paper is a preliminary attempt to re-frame our (broadly cast) collective research agenda in the wake of the digital communication age.

In a recent article which surveys the two seemingly definitive areas of “Information Science” (exemplified by the Journal of the American Society for Information Science and Technology) and “Information Systems” (exemplified by the Journal of Information Systems Research), Sawyer and Huang (2007) put forth the following:

We speculate that scholars in [information systems research] need to be more explicit about its basic constructs (even to the simple level of articulating the range of possible ways to characterize these basic constructs) to more easily connect to, and to be understood by, other disciplines (Sawyer & Huang 2007: 1444).

Sawyer and Huang (2007) seek to find common ground among researchers within the fields of Information Science and Information Systems. Their purpose is to gain some insight into how to increase collaboration both within and between fields. Additionally, they are interested in understanding how to build bridges between these fields and others. We believe that their interest in these collaborating behaviors actually illustrates the “basic constructs” they seek. We do not view this as a coincidence given they are “IS” researchers who have positioned themselves within the larger “i-field,” which we view as being composed of—at least—Information Science, Information Systems and Library Science researchers. We posit that those who view themselves as being tied to the i-field share a common interest in collaborating behaviors and technologies that support these behaviors (see how we define our broad use of the term “collaborating” below) and that these interests converge on phenomena associated with collaborating and communicating behaviors that characterize the field. It seems that it is an interest in these phenomena and these technologies from which all inquiries into the IT artifact, IT per se, individual, social or organizational processes, representation, retrieval, transfer, etc. emerge. We suggest that the increase in the salience of these characterizing phenomena are a direct result of employing technology to communicate and collaborate in our increasingly intertwined global environment.

Further, we suggest that one of the profound impacts of IT is not derived merely from its wide scale adoption, but from the resultant increase in salience of collaborating behaviors embedded within every social environment. For example, we are now able to point directly to a set of artifacts (e.g., e-mails, IM messages, blogs, etc.) written by various individuals who are collaborating to solve a specific problem and validly view it as evidence of a “team” or “community.” Yet, not a single face-to-face interaction among their members is required before doing so. Those within the i-field are currently positioned to explain why this characterization is valid and why it can be reliably made across all contexts in which humans interact. Most importantly, this brings the field beyond explaining how or why this type of human interaction is possible—explanations that are seemingly self-evident to researchers in any field of research, practitioners from any area of business as well as consumer of personal computing devices world-wide. We believe that IT has worked to illuminate who we are as humans and that the emergence of the i-field is a direct result of a common interest in this. It is insight into who we are (as a field) that will:

- (1) create the gravitating effect discussed by Sawyer and Huang (2007),
- (2) uniquely define what it means to be a researcher in the i-field to those in other fields as well as the world, and
- (3) lead to better design of systems and technology which will in turn enable those in the field to have a more profound impact on the world (let alone other disciplines).

The Internet, and more specifically the Web, has changed the functional definition of community by illustrating the collapse of the geographical/spatial constraint on communicating. Community implies an organizing principle more intimately tied to the problem/situation than to person which has resulted in communities that are more narrowly focused in the range of problems/situations that define them. The Web has also spawned a multitude of types of artifacts (what our field calls “information”) that differ in, as yet, unpredictable forms, units, etc. that our inherited logic (so far at least) is incapable of “managing,” e.g., voice mail artifacts, texting artifacts, discussion group messages, FAQs, discussions, etc. There is an inconceivably immense and rapid increase in the total volume of these types of artifacts. Another way to describe the Internet and its multi-media offspring, the Web, is the removal of many hierarchical or top-down constraints, both in terms of the “means of production” (à la Marx) and in terms of centralized organizational/political control. These examples work to illustrate why we believe a more focused and solidified agenda for the i-field is not only timely, but—most importantly—necessary.

In this paper we put forth the ideas expressed above as well as our own research agenda with the hopes of sparking a more convergent debate about the i-field and interests of those in it. We put forth our agenda as an example of one that is:

1. geared towards gaining insight into collaborating behaviors—albeit from our interdisciplinary perspective (by definition, given our linguistic and cognitive backgrounds)—that can be folded into, stand in firm contradiction to or extended by other researcher agendas,
2. defined and bounded in such a way that we can delineate it from research agendas that work to characterize other fields,
3. positioned to be levered for the design of systems geared towards complementing, or potentially enhancing, collaborating behaviors.

We begin with a brief discussion of our perspective on what it is that yokes researchers within the i-field. We then define terms that are relative to our own research agenda, but seemingly relevant to a number of others as well. We then discuss how our agenda has and continues to unfold over time and how it relates back to the i-field-specific themes discussed at the beginning of the paper. Given our discussion above about the field, we continue with some discussion on the nature of our inquiry in collaborative behaviors to give some sense of the flavor of our discussion (we save all assumptions and definitions for the actual paper).

Our field has dedicated a great deal of time and energy examining content and its technology in the last fifteen years, but we believe that we must commit significant resources into understanding human beings and their behavior from a more functional and purer user-based perspective. We are certainly not alone in this view (e.g., Dervin & Nilan, 1986; Saracevic, 1997; 1999; Sawyer & Huang, 2007).

To the extent to which we can find patterns (i.e., similar actions and sequences of actions, rationales, values, descriptions, behaviors, etc.) in the way people collaborate and communicate about their problems/situations, we can examine these patterns to see if they can be employed effectively for the organizing and subsequent retrieval of artifacts. There is some evidence that indeed, they can be so employed (e.g., Nilan, 1992; Yoon & Nilan, 1999; Nilan & Mundkur, 2007). Note, that “problem/situation” is a contextual anchor to which all behaviors we are interested in exploring are tied. This allows us to observe behavioral regularities in a systematic and relatable manner. We argue that these same patterns may be suitable for organizing non-traditional resources that our current approaches to the organizing/retrieval dilemma do not accommodate (e.g., maps, images, people, places, etc.).

Further, we believe that our functional understanding of these patterns will serve to inform our design of systems and technology so that we can facilitate the organizing and retrieving of artifacts as those artifacts are generated. For example, e-mail currently has only “To,” “From” and “Subject” fields. We believe that, informed by our understanding of the collaborating and communicating “habitus” of human beings, our discipline could add additional fields (with a minimum of explanation to users) which would serve subsequent organizing and retrieving.

Two courses of action are suggested: One, to generate rich descriptions of the community-based communicating behaviors and subsequent cognitive behaviors (i.e., perceptions and step taking) of the individual community members to search for patterns. Two, based on these rich descriptions, apply our experience/“habitus” proactively to facilitate the direct communicating behaviors while facilitating the effective and efficient use (and re-use) of vicarious experience.

We conclude our paper with specific user-based research areas into collaborating and communicating behavior linked to system design.

References

- Bourdieu, P. (1999). The social conditions of the international circulation of ideas. Richard Shusterman (Ed), Bourdieu; a critical reade, Oxford: Blackwell, 220-228.
- Dervin, B., Nilan, M. S. (1986). Information needs and uses, *Annual Review of Information Science and Technology*, 21, 3-33.
- Nilan, M. S. (1992). Cognitive space: Using virtual reality for large information resource management problems. *Journal of Communication*, 42:4, 115-135.
- Nilan, M.S., Mundkur, A., Structuring virtuality: A user-based example for e-commerce. K. Crowston, S. Sieber, E. Winn (Eds), *International Federation of Information Processing*, Vol. 236, *Virtuality and Virtualization*, Boston: Springer, 53-66.
- Saracevic, T. (1997). Users lost: Reflections on the past, future, and limits of information science. *ACM SIGIR Forum*, 31:2, 16-27.
- Saracevic, T. (1999). Information science. *Journal of Information Science & Technology*, 50:12, 1051-1063.
- Sawyer, S., Huang, H. (2007). Conceptualizing information, technology, and people: Comparing information science and information systems literatures. *Journal of Information Science & Technology*, 58:10, 1436-1447.
- Yoon, K., Nilan, M.S. (1999). Toward a reconceptualization of information seeking research: Focus on the exchange of meaning. *Information Processing & Management*, 35, 871-890.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

48.1 In the Spirit of Collaborating

by
M. S. Nilan and M. A. D'Eredita
i-School
Syracuse University

Abstract

A field is judged by the questions it addresses. We believe that the i-field is at a critical juncture that is defined by increasing standards for profound, value-added insight as technology becomes more ubiquitous. While we see value in focusing on the "I" more than the "IT" in IT, we believe that profound insight will only come through a deeper understanding of the intersection among human beings, information and technology. We see the basic goal of our field as helping to facilitate interactions among individuals, groups and "information" through more informed design of technology. We assume that human beings are fundamentally social beings and both information and technology to be artifacts of human experience proffered in the spirit of collaborating. We begin with a discussion of collaborating and communicating behaviors which are inherently linguistic (broadly cast) and - we believe, by definition - functional in nature. This position is logically prior to artifacts or any technology associated with the creation, dissemination, storage, retrieval or other processing of "information." We argue, for example, that it is insufficient for our field to bound a domain after artifacts are created - we must become engaged in the creation of artifacts, specifying how the artifact will be useful to the user (either at present or in the future), and for which problems/situations the artifact is pertinent. If one assumes an active user focused on addressing their own needs as well as those of others--as opposed to a more passive stance generally made by many of the traditional behavioral fields--then insight into artifact creation, use and re-use will not only result in profound insights into the human condition, but more informed system design. We humbly invite you to collaborate with us in this endeavor. We feel that it is not only critical to intensify this discussion, but both timely and necessary.

Introduction

In the twelve years since the introduction of the World Wide Web (April, 1995), we have seen dramatic changes in technology that have allowed for a plethora of collaborative environments and new means for communicating to emerge. This, in turn, has

generated a plethora of new artifacts that we call “information.” These technological artifacts are a direct result of specific and ubiquitous human behaviors manifesting themselves in seemingly distinctive ways. These changes increased the salience of a set of formidable and—we argue—distinguishing phenomena our field is now addressing. This paper is a preliminary attempt to re-frame the discussion about our (broadly cast) collective research agenda around these phenomena.

Vannevar Bush (1945) posited the ultimate information retrieval mechanism, which he called the “memex.” Memex organized content in a collection based on “associative indexing.” In his view, “associative indexing” would permit retrieval based on association and context rather than strict categorical indexing. Bush’s efforts have been seen as the conceptual framework for subsequent work on hypertext, which is the central data metaphor for the digital age. We view this as a good example of a technological focus absent insights into human or user behavior. For example, Bush did not specify the nature of the associations among nodes in the hypertext. We believe that functional associations between what information is about and how the information might help a user is a logical source of insight into Bush’s associations. In other words, we believe that the source of the association that Bush alluded to is at the intersection of the information artifact and its benefit, or derived use. The purpose of this paper, however, is not to specifically criticize or extend Bush’s ideas or to recast general ideas about a research agenda that is more about the “I” than the “IT” in IT. Rather, we wish to raise the question of whether or not moving the focus to the “I” is going far enough to define our field in such a way that it will lend insight into fundamental issues like those brought up by Bush over 60 ago.

Our intent is to explore the beginnings of a coherent conceptual framework that is derived from a fundamental understanding of what leads human beings to collectively create, use and re-use our experience when addressing life’s challenges. We assume that these collaborating behaviors are fundamental to being human and are exhibited by people communicating experiences and understandings with each other. These fundamental behaviors inevitably result in constructed artifacts that are necessarily constrained by the functionality of technologies designed to generate, access and disseminate information. We believe that the i-field is uniquely positioned and capable of gaining profound insight into these fundamental behaviors only if we adopt a truly interdisciplinary stance that incorporates the “I,” the “IT” and the “H” (i.e., the human being) when trying to understand the whys and hows of creating, using and re-using information.

In a recent article which surveys the two seemingly definitive areas of “Information Science” (exemplified by the Journal of the American Society for Information Science and Technology) and “Information Systems” (exemplified by the Journal of Information Systems Research), Sawyer and Huang (2007) put forth the following:

We speculate that scholars in [information systems research] need to be more explicit about its basic constructs (even to the simple level of articulating the range of possible ways to characterize these basic constructs) to more easily connect to, and to be understood by, other disciplines (Sawyer & Huang 2007: 1444).

Sawyer and Huang (2007) seek to find common ground among researchers within the fields of Information Science and Information Systems. Their purpose is to gain some insight into how scholars might increase collaboration both within and between fields. Additionally, they are interested in better understanding how to build bridges between these fields and others. We believe that their interest in gaining insight into these collaborating behaviors actually illustrates the “basic constructs” they seek. We do not view this as a coincidence given they are “IS” researchers who have positioned themselves within the larger “i-field,” which we view as being composed of—at least—Information Science, Information Systems and Library Science researchers. We posit that those who view themselves as being tied to the i-field share a common interest in collaborating behaviors and technologies that support these behaviors (see how we define our broad use of the term “collaborating” below) and that these interests converge on phenomena associated with these behaviors that characterize the field. It seems that it is an interest in these phenomena and these technologies from which all inquiries into the IT artifact, IT per se, individual, social or organizational processes, representation, retrieval, transfer, etc. emerge. We suggest that the increase in the salience of these characterizing phenomena are a direct result of using technology to communicate and collaborate at all levels of an increasingly intertwined global environment.

Further, we suggest that one of the profound impacts of IT is not derived merely from its wide scale adoption, but from the resultant increase in salience of collaborating behaviors embedded within every social environment. For example, we are now able to point directly to a set of artifacts (e.g., e-mails, IM messages, blogs, etc.) written by various individuals who are collaborating to solve a specific problem and validly view it as evidence of a “team” or “community.” Yet, not a single face-to-face interaction among their members is required before doing so. Those within the i-field are uniquely positioned to explain why this characterization is valid and why it can be reliably made across all contexts in which humans interact. Most importantly, this brings the field beyond explaining how or why technology enables this type of human interaction—explanations that are seemingly self-evident to many simply because IT is becoming embedded in peoples’ daily lives and business process. In other

words, the fascination with technology is fading and potentially turning more towards an expectation of more connectivity. Beyond the “I” and certainly the “T,” we believe that IT has worked to illuminate who we are as humans and that the emergence of the i-field is a direct result of a common interest in this. It is insight into who we are (as a field) that will:

- (4) create the gravitating effect discussed by Sawyer and Huang (2007),
- (5) uniquely define what it means to be a researcher in the i-field to those in other fields as well as the world, and
- (6) lead to better design of systems and technology (e.g., organizing artifacts according to Bush’s (1945) notion of “associative indexing”) which will in turn enable those in the field to have a more profound impact on the world (let alone other disciplines).

Our field has dedicated a great deal of time and energy examining content and its technology in the last fifteen years, but we believe that we must commit significant resources into understanding human beings and their behavior from a more functional and purer user-based perspective. This perspective differs from those of many other fields focused solely on behavior—and potentially viewed as being focused on the user as well (e.g., those within the psychological, economical and political science fields)—by viewing the user as an active and creative participant who is actually working towards constructing their environment as opposed to a more passive participant merely reacting to their environment. We are certainly not alone in this view (e.g., Dervin & Nilan, 1986; Saracevic, 1997; 1999; Sawyer & Huang, 2007).

There are many ways to re-frame our research agenda. One is to conceptualize the artifacts that we call information as byproducts of various attempts by human beings to collaborate and share experiences with each other. Some already under way include applying pre-digital solutions to digital content (e.g., meta language classification of web pages), others are uniquely digital solutions (e.g., natural language processing for “filtering”) and still others are a blend of the pre-digital and digital (e.g., folksonomies). We offer a perspective that has emerged out of a juxtaposition of the so-called “user-based,” cognitive and social cognitive approaches (Nilan & D’Eredita, 2005; D’Eredita & Nilan, 2007) and believe that it also works to illustrate the focus of inquiry discussed above.

What follows is strictly an illustration of how we frame our research within the i-field as discussed above. It is important to note that we do not view the perspective we put forth below as exemplary of the i-field. We believe that the common ground we share with all i-field researchers is in our overarching assumption of an active user and our interest in collaborating behaviors and how to realize the potential of IT to support these behaviors. It is from our interest in collaborating behaviors that we believe we can begin a dialogue with any researcher considered to be part of the i-field as well as those who are not. Below, we put forth the foundation of our research on collaborating behaviors. We differentiate ourselves from other researchers within the field by relating information to human collaborating about specific situations or problems (discussed below). It is also important to note, that our anchoring curiosity in specific types of collaborating behaviors allows us (a linguist and cognitive psychologist by “trade”) to work in a complementary and fulfilling manner. From our perspective, our “home disciplines” only add perspective to the more important common phenomena in which we are interested in gaining insight.

It is in the spirit of collaborating that we present this paper. Our primary motivation is to see if others are able to frame their research within a field focused on all aspects of human collaborating behaviors. We begin by defining key terms relevant to our own research agenda, but seemingly relevant to a number of others as well. We then discuss how our agenda has and continues to unfold over time and how it relates back to the i-field-specific themes discussed at the beginning of the paper. Given our discussion above about the field, we continue with some discussion on the nature of our inquiry into collaborative behaviors in order to give a sense of the flavor of our discussion.

Our Anchoring Definitions

49. “Collaborating” is the social coordination (largely communicative) through which human beings observe their environment and solve problems/address their life situations. Conceptually, Carter’s (1980) “social Darwinian” description of human existence argues that humans have an observing and a moving capability. Those individuals who observe and move effectively will prosper. Through collaborating, individuals can substantially improve their individual observing and moving capabilities. Two heads are better than one; together we are stronger. It is through linguistic communicating this basic social stance is manifest, i.e., we inform one another to our mutual benefit. The artifacts we generate in this communicative process, however distorted by structure (e.g., publication or broadcasting constraints) is what we have been calling “information” (further defined below).

“User-based” refers to an epistemic position (i.e., what can be known about being human) that validates the reality of human beings as they perceive it as a foundation for the design of systems (e.g., a series of steps intended to solve a human

problem/situation/ context which may or may not involve reification of some or all of the steps as “technology”). There is currently a great deal of misunderstanding about user-based where some believe that we need to understand the users per se (i.e., the so-called “individual differences” approaches) in order to fulfill our research agenda. However, we believe we need to focus on the users only within the situations/problems that work to both motivate and engage human collaborative and communicative behaviors and in some cases, only the situations/problems in which users are engaged.

“Problem” (Taylor, 1986; Nilan, 1992) or “situation” (Dervin, 1983; 2001; Dervin & Nilan, 1986) is the time/space context for human cognitive behavior. These are very broad terms, not to be confused with the much narrower “problem solving” notions from cognitive psychology, management, etc. In essence, the terms refer to a “chunk” of time/space about which people collaborate/communicate and ultimately the focus of system design – helping users share and create meaning to address their situations/solve their problems through access to resources (information/data; computing functionalities; links to others, e.g., experts).

“Language” is the means through which human beings create and exchange meaning. This is accomplished through a series of co-focusing, co-orienting, and collaborating behaviors (Kim, 2003) which serve to orient one person’s context to the other’s. Following the so-called “Prague Functionalists,” a unit of meaning has two necessary components in this orienting process: “Topic,” or what I am talking about and “Comment,” or how the topic at hand “fits” my context (i.e., problem/situation and any associated goals or envisioned end states) (Jakobson, 1963). Meaning is therefore, not exchanged per se but is converged upon over an iterated series of utterances over time. Language as we are defining the concept has nothing to do with the difference between Chinese Mandarin and American English. Rather, language is an extremely broad concept that encompasses all manner of expressing relationship, existence and movement and it is much more than merely a tool. Language is employed to orient other people, one to the other (whether in a disclosure sense or in a deceptive sense), orient other people to a specific context (time/space bound problem/ situation, i.e., topic), and/or to specify individual experience with that context in terms of values, perceptions of situational conditions, goals or end states (i.e., comment).

“Information” is a resource which (we assume) helps the human(s) involved in addressing a problem/situation facilitate progress towards her/his/their desired goal or end state. [Note: “facilitate” implies constraint to perception of the environment and/or perception of movement possibilities and consequences. A resource that facilitates may not necessarily be empowering, it may also be inhibiting while a constraint may not necessarily be inhibiting, it may also help to facilitate “cognitive movement” (defined below)]. All information is bound by the context in which the collaborating occurs. For example, a textbook by an expert in Alzheimer’s disease may contain insight to help me deal with a friend with the disease but it is buried in a book which was structurally constrained to fit a 15-week educational program. Since the small bit of information that might help me is not part of an actual conversation between me and the author, it is extremely difficult for me to find – there are no co-focusing or co-orienting features in the text linking me, the reader, to the writer and scarce opportunity for this to ever occur. Even when we observe two people in conversation as they collaborate on one or the other’s situation/problem, there is much that is not articulated (because it is already understood) that the outside observer has no access to which renders any technological recording of the interaction necessarily incomplete. Similarly, technological artifacts are necessarily incomplete expressions of context and, therefore, inadequate representations of the problem/situation from the perspective of either person involved. This in turn presents a difficulty when we seek to access and re-use the artifacts, especially when they are “located” in a huge “pile” of other inadequate representations. Note that we can posit a continuum of information artifacts, from an artifact addressing how to make a chocolate cake (for example) where there may be many possible contexts where a chocolate cake is appropriate, to an artifact addressing the nature of Alzheimer’s disease where the help needed by a patient is likely to be different than the help needed by a relative of the patient. In neither case is the user’s context likely to be present in the information artifact itself—albeit a seemingly more profound issue in the case of Alzheimer’s disease.

“Cognitive movement” is a central metaphor in Dervin’s (1983) Sense-Making which is employed to describe the human experience of being alive over time. In this metaphor, experiencing life is described as taking steps (mentally as well as perhaps physically) through problems/situations, sometimes concurrently (although not simultaneously) or serially. Cognitive movement invariably implies emotive steps. Cognitive movement is one of two central concepts in the user-based approach we put forth in this paper. It is inseparably tied to the concept of problem/situation.

“Uncertainty” (Carter, 1980) refers to the aspects of an individual’s (or individuals’) perceptions of a problem/situation that are not already pre-determined. Since change (in self, others, environment and problem/situation) is pervasive, uncertainty is always (except with reflexive and habitual behavior) an aspect of human perception. Dervin (1983) refers to a “gap” to represent the uncertainty inherent in an individual’s perception of her/his position relative to that individual’s movement through the situation/problem/ context at hand. When uncertainty is perceived by the individual, s/he can attempt to resolve the uncertainty

by sharing meaning or creating meaning through collaborating with others (either directly through conversation or indirectly through artifacts of past conversations).

From a monadic perspective, “meaning” describes the sense that an individual makes of aspects of the environment inherent in her perception of her problem/situation in terms of her understanding of the environment and/or her movement relative to that environment to reach her goal or end state. This meaning is based on past experience and certainly based on past and present conversations with others (either direct or vicarious). In this way, meaning can be seen as the creation of new “sense” (Dervin, 1983) about “how to bridge the gap” (i.e., addressing uncertainty), it can also be seen as the creative re-interpretation of old meaning to “fit” new situational conditions. Although meaning is created directly by experiencing a problem/situation, experience can also be vicarious – experienced indirectly through observing others. A dyadic (or larger) model of sense making can help to guide our efforts because the observing of others’ experiences and behaviors is realized through conversing with others about the specific context given specific “comments.” Such “comments” are rare in most formal artifacts but comprise the bulk of the verbiage in a conversation (see Yoon & Nilan, 1999). This has become even more evident in our digital, post Web world. From birth, individuals are enmeshed in a collaborative network of others who attempt to help the individuals observe and move effectively through collaborative communicating. All communicating is through language grounded in individual experience (direct or vicarious) and meaning is created by the similarities and differences among individual experiences. Much meaning (some, i.e., structural views – e.g., Giddens (1991), Bordieu (1999) – would say “all” meaning) is perpetuated over time through pervasive structuralization (e.g., historical, cultural) constraint and extended experience with a situation/problem (or group of related situations/problems). [Note: “communicating” does not necessarily result in the rare phenomenon of “communication” (Carter, 1980). So communicating only implies the attempt to share experience and orientation, not the result]. One point of the foregoing is that explicit comments in an artifact (which is much more likely in a conversation than a formal publishing or broadcasting artifact) allow a subsequent user to quickly determine whether the artifact will be of any use to him/her.

“Knowledge” refers to experience with a specific problem/situation or set of related problems/situations (D’Eredita and Barreto, 2006). We assume the following in regard to knowledge (D’Eredita and Barreto, 2006: 1824):

1. The nature of knowledge is episodic, thus personal, situation-/problem-specific and acquired through experience, i.e., there is no abstract or non-contextualized product “stored” somewhere in the brain.
2. Experience results from the construction and relating of episodes, but does not necessarily result in specifically “intended” knowledge.
3. The proliferation of experience by individuals within a social context (e.g., dyad, group, team, organization, community, etc.) is the result of constructive and collaborative communicating behaviors by which two or more individuals collectively focus attention, thus collectively construct relatable episodes—from the perspective of both the intra- and interpersonal.

Logan (1988) defines an episode as follows:

Processing episodes consist of the goal the subject was trying to attain, the stimuli encountered in pursuit of that goal, the interpretation given to the stimuli with respect to the goal, and the response made to the stimulus (Logan 1988: 495).

To paraphrase Logan’s quote in Dervin’s (1983) terms, making sense of a situation involves the goal or end state that the individual has in mind, the situational conditions as perceived and interpreted by the individual as s/he moves through the situation, and the cognitive behaviors (especially those that serve to collaborate and to communicate with others) that the individual takes to move towards his/her goal. This entire “chunk” of time and space (i.e., episode) is the context which is the central focus of user-based research put forth in this paper. That we do this by collaborating and communicating (via language) with others implies that the constructed “meaning” is tied to social cognitive behaviors among individuals.

This is very similar to sense-making from the perspective of Carl Weick (albeit more specific):

The process of sensemaking is intended to include the construction and bracketing of the text-like cues that are interpreted, as well as the revision of those interpretations based on action and its consequences. Sensemaking is about authoring as well as interpretation, creation as well as discovery (Weick 1995: 8).

Overlap among episodes allows for a stream of discrete episodes to result in a seemingly fluid flow of experience. We would not be able to make sense of our experiences if this was not true (Weick, 1995). For example, D'Eredita and Barreto (2006: 1829) suggest the following:

1. A new episode might be associated (individually and collectively) with previous episodes because of similar qualities as defined above. That there is some type of incongruity results in the most obvious need for sensemaking (Weick, 1995). It presents the need for constructing a potentially unique episode that is congruent enough for it to be associated with past episodes; or, for it to make sense. This is not an abstraction, but a direct comparison made through human-cognitive behaviors. Episodes are explicitly related or not explicitly related in some manner. In the latter case, previous episodes, however, most likely continue to indirectly influence behavior whether individuals are aware of this influence or not (Reder, 1996).
2. On the other hand, a new episode might reinforce a previous experience because of similar qualities. Here, one can assume a high probability that newly constructed episodes will be related to past episodes. Unlike the previous case, the amount of effort required to make sense of an episode is minimal, thus allowing for specific stimulus-response behavior to be optimally reinforced (Wyer, 1997).
3. A new episode might also fall between the two extremes. This is perhaps the gray area in which one is said to rely upon previous experience with similar episodes/ situations/problems referenced by Nonaka and Takeuchi (1995) in their view of tacit knowledge. Here, one is simply afforded the chance to do less sensemaking than in the first case presented while also being able to leverage an array of previous episodes that are functionally (or not) similar in quality.

Note that Bordieu's (1999) notion of "habitus" can help us understand scenarios two and three above. Habitus refers to familiarity with a particular situation/problem (through experience) such that the resources of the individual are more oriented towards the present uncertainty of situational conditions that are out of the ordinary (e.g., have changed since the last time the person has addressed the situation/problem). This is not to be confused with "habit" which is a reflexive behavior (non-reflective) (Cohen, 2007).

To juxtapose all of the above with ideas presented by D'Eredita and Barreto (2006):

A robust cognitive and social cognitive set of behaviors emerges if we assume experience is episodic and is the construction of and relationship among episodes. This allows individuals to be practical and "reliable" (Weick & Roberts, 1993) in regard to behavior ... Episodes may be constructed given a specific problem, but not necessarily in a manner that makes them representative of an array of potential contingencies (D'Eredita and Barreto, 2006: 1829).

This leads us to assume the following in regard to episodes, ergo experience (see D'Eredita and Barreto, 2006, for a more complete discussion on these assumptions which are a juxtaposition of assumptions from a perspective representative of the field of cognitive psychology, particularly that of Logan (1988), with that of an i-field perspective):

1. The constructing and relating of experience is episodic: Each episode is associated with one or more previous episodes, but is constructed as a separate co-occurrence.
2. Constructing new episodes is obligatory upon attention: That which is attended to is automatically perceived and embedded within the context of a given episode.
3. Relating to previous episodes is obligatory upon attention: Previous experience directly impacts behavior if a current situational condition is attended to and is associated with previous episodes in some way.

"Memory" refers to the ability of human beings to associate present situational conditions with perceptions of past situational conditions relative to the current episode. The associating is not exact, so we see limitations in the objectiveness of human memory and we certainly should consider the individual's criteria employed for associating as inexact. The extent to which the associating is based on "habitus" or inherited criteria and logic represents increased likelihood that associations made for evolving uncertainty in situational conditions in the present and future will not be up to the task. It is important to note that while our perspective is cognitively slanted, our notion of episode (given its relationship to problem/situation) applies to

context and both individual and social behavior. It is at this point in which we—if we have yet to do so—radically diverge from traditional cognitive theory.

We further diverge by suggesting that any notion of “storage” in relation to “memory” is misleading. Our epistemological stance places the individual in the present and notions of “past” and “future”—thus time—as cognitive constructs that facilitate sense-making from a functional perspective. Simply put, we are not comfortable assuming the world is as it “seems” to be, but rather how the world seems to be is functionally similar to “how it is.” We assume that the nexus of the cognitive behaviors described above is that of a neurological functioning organ—the brain. The physical foundation from which more abstract cognizing emerges is assumed to be in a constant state of change as it interacts with the world. It is because of this that we see no need to accept notions of “memory” like that adopted from a (computer-based) information-processing perspective. While we assume experience is “bounded” in the episodic manner discussed above and that each episode is processed as separate co-occurrence, we do not assume that these episodes are “stored” in this manner waiting to be “retrieved” or “recalled” at a latter time. We view behavior as resulting from cognitive behaviors that have been shaped by previous experience and as a response to a given problem given its (and an individual’s and/or group’s) state at a specific point in time and space.

Further, we do not assume behavior to be deterministic, albeit potentially implied given what has been said thus far. We view behavior as being tied to experience in a probabilistic manner and to current constraints imposed by one’s interpretation of the problem (which is also relatively uncertain and probabilistic in nature). More experience implies that the individual or group has a better chance of usefully, or functionally choosing, or creating steps that facilitate individual or group progress towards the desired goal or end state. “Expert,” as its etymological roots imply, refers to one who has experience. Culturally we have cast experts as the ones who can most effectively and efficiently respond to a specific problem by choosing or creating appropriate steps. However, experience is based on past uncertainty rather than present or future uncertainty. This past orientation may potentially limit the expert in dealing effectively and efficiently with emerging new situational conditions.

“Creativity” refers to an individual perceiving or taking steps that are beyond his/her existing (direct or vicariously “inherited”) experience with the problem/situation at hand. As such, it is different from a judgment made externally, e.g., a “creative” work. Given the necessary association with the problem/situation at hand, we view creativity as being best conceptualized as a time/space bound cognitive behavior rather than a personality-level characteristic of the individual. The issue for researchers and managers thus changes from identifying “creative individuals” to identifying creative perceiving or step taking given the current collective experience available (e.g., Weisberg, 1993).

“Community” refers to the (direct and vicarious) communicating among individuals engaging with or involved in addressing uncertainty in one or more problems/situations. Humans are social creatures whose development begins (usually) in “families” and is subsequently nurtured in extra-familial sub-communities (e.g., schools, religious activities, etc.) which serve to broaden the individual human’s perspectives from an existential self-centeredness towards (the inclusion at least) a more social- or network-

centric focus. Community is defined by the range of situations/problems with which the collective is engaged. “Organization” differentiates itself from a more emerging community in that it has already organized—albeit in some cases continuously adapting in order—to address a distinctive set of problems. In this sense, the more formal structure of an organization becomes more of a constraint (as discussed above) when compared to the more informal structure of a community. This is not meant to imply that community or organization are simply less or more developed forms of the same social entity. However, we do mean to imply some overlap in that they are a direct—albeit not exclusive—result of fundamental collaborative behaviors yoked by a set of common problems.

Below, we discuss how our agenda has and continues to unfold over time and how it relates back to the i-field-specific themes discussed at the beginning of the paper.

A Research Agenda Characteristic of the i-field

Constraints are logically tied to the problem/situation at hand and the more micro situational conditions that serve to contextualize the communicating behaviors which includes any potential IT (D’Eredita & Nilan, 2007). If we as researchers move from an individual-as-unit perspective to something that includes a more time/space bound problem-as-unit as well as the embedded, and even more micro, cognitive step-as-unit perspective, we might make some headway in understanding and managing the changes that impact our field’s objective of providing access to resources. {Note: there is an even more micro unit of analysis common in the Sense-Making research of Dervin and her colleagues (e.g., Dervin, 2003) which is the “gap” which facilitates inquiry into very specific information and resource use, resource evaluation, etc. A “gap” is usually operationalized as questions that a respondent has in her/his mind at a particular cognitive step. A separate unit of analysis is needed for this fine-grain issue because multiple gaps are possible at any specific step}.

As a field, we seek to create or discover patterns that serve to guide our organization of resources which in turn, must guide information seekers in locating resources for subsequent re-use. We argue that patterns which are derived from the technological artifacts as we find them are ineffective in facilitating either appropriate levels of agreement in classification of artifacts or retrieval by users (who we know won’t “read the manual”). The technology that produces the artifacts provides no insight into the logic of this organizing/retrieving endeavor whatsoever, it is a human behavior that ultimately dictates the function of any organizational scheme. Therefore, to the extent to which we can find patterns (i.e., similar actions and sequences of actions, rationales, values, descriptions, behaviors, etc.) in the way people collaborate and communicate about their problems/situations, we can examine these patterns to see if they can be employed effectively for the organizing and subsequent retrieval of artifacts. There is some evidence that indeed, they can be so employed (e.g., Nilan, 1992; Yoon & Nilan, 1999; Nilan & Mundkur, 2007). We argue that these same patterns may be suitable for organizing non-traditional resources that our current approaches to the organizing/retrieval dilemma do not accommodate (e.g., maps, images, people, places, etc.). After all, these patterns represent the universal “habitus” of human survival through collaborating and communicating which is far older than the mere 4,000 years since the invention of writing. And the patterns (e.g., between “topic” and “comment” or between one step and the next step) represent associations readily interpretable to users.

Further, we believe that our functional understanding of these patterns will serve to inform our disciplinary design of systems and technology so that we can facilitate the organizing and retrieving of artifacts as those artifacts are generated. Note that constraints associated with design should be viewed as necessarily balanced between enabling natural—and potentially idiosyncratic—human behavior and guidelines for effective creation and exchange of meaning through “information.” For example, e-mail currently has only “To,” “From” and “Subject” fields. As Yoon & Nilan (1999) suggested, these might be augmented with “comment” fields such as ‘how the email will help the recipient understand his/her problem better’ and or ‘how the email will help the recipient figure out what to do at this time’ and/or ‘why this email is important to the recipient.’ The focus in this example is on “comment” and therefore the functional associations between users’ views of reality and their movement through that reality. This kind of association – between “topic” and “comment” – represents one fairly concrete association (of potentially many possible) that might enable our field to actually guide the generation of artifacts that incorporate Bush’s “associative indexing” as the artifacts are generated. We believe that in this way our discipline could lead insightful changes to relatively ubiquitous technologies that would serve subsequent organizing and retrieving. Again, this relatively small change would be based on insight into the fundamental collaborating behaviors discussed above and not the technology per se.

49.1.1 Turning an Agenda into Action

While we emphasize creation, use and re-use of experiences above, here we will focus on only the latter two, but with the assumption that all that is discussed below is directly applicable to the design of systems specifically focused on the creating of

information. Two courses of action are suggested in regard to use and re-use by the foregoing: The first is to generate rich descriptions of the community-based communicating behaviors and subsequent cognitive behaviors (i.e., perceptions and step taking) of the individual community members to search for patterns. The second is, based on these rich descriptions, apply our field-specific experience/"habitus" proactively to facilitate the direct communicating behaviors and at the same time, facilitate the effective and efficient use (and re-use) of vicarious experience. Since systems (including classification/ organization of artifacts) have been defined above as "a series of steps designed to solve a human problem," empirically derived patterns will inform us as to the selection of appropriate steps and sequences of steps. We see these patterns as superlative examples of associative dimensions alluded to by Bush (1945).

Based on this change in focus, the following user-based and problem/situation-based research agenda is proposed for the study of collaborative behaviors within communities, but with the assumption that a similar approach can be applied to, for example, teams:

First we would seek to describe behaviors within the collaborative communities that punctuate human lives (intra-community). Rich descriptions of the communicating and collaborating behaviors (not merely perusal of existing technological artifacts which are, by definition, meaning-challenged in-and-of-themselves) and their rationales (i.e., both topic and comment) can be examined for evidence of patterns in: (1) effective and/or efficient behaviors (steps); (2) sets of behaviors, e.g., roles (but NOT as characteristics of individuals) as functional sequences; and (3) reward structures and constraints that facilitate communicating and therefore, collaborating. Part of our disciplinary confusion vis-à-vis these behaviors are like a "fish in water" scenario – we are embedded in social interaction and therefore find it quite difficult to posit the behaviors as an object of inquiry. We believe it to be an imperative initial focus of inquiry. The flexibility of notion of "community" is particularly useful here because it would allow researchers to tailor their observations to groups with limited ranges of problems/situations (e.g., specific organizations or organizations focused on providing similar products or services). This in turn would allow us to pursue a basic research agenda into our primary phenomena and provide concrete practical insight into the targeted communities. As we pursue our rich descriptions, "community" as a unit of analysis will facilitate many analyses across different communities to get at the essence of human communicating and collaborating as well as experiment with various artifact creation possibilities.

We can subsequently examine similar phenomena in the communicating among communities. Of particular interest here is referrals from one community discussion to another – when, under what conditions, etc. given the "horizontal" communication engendered by the Web. We believe that insight into this specific subset collaborative behaviors will be particularly useful for many information seeking scenarios in the digital age.

Once we have adequately described these phenomena and looked for relationships among behaviors and various use criteria, our collective understanding of the human perquisites inherent in creating, using and re-using others' experiences we can then become proactively engaged in research and design of the actual production of artifacts which incorporate these efficiencies to facilitate appropriate vicarious re-use of the documented experience (e.g., adding fields to email/voicemail/texting interfaces). This will, in turn, facilitate subsequent re-use of the technological artifacts in more of a conversational manner as opposed to the current topical "keyword" strategy. We believe this kind of search interface (e.g., specify "topic" and then specify "comments" [plural employed here is deliberate]) will be more natural for the large majority of users – with no manual or knowledge of Boolean logic necessary. In essence, this is what Yoon & Nilan (1999) suggested - the organization of resources (both communicative and artifactual) within and between problem/situation-based communities. Over the (currently unimaginably huge) hypertext data store of artifacts that tomorrow's Web will contain, we believe that this communicative/collaborative framework will facilitate the sharing of resources within and among communities, and ultimately, facilitate the effective/efficient access of individuals to the communities which best "fit" their individual problems/situations.

Whether or not you subscribe to Bush's (1945) "memex" vision, the potential power of his "associational indexing" to manage inherently unmanageably huge data set seems logically possible if we posit the associating as an extension of how human beings help each other make sense of and move through the sequence of situations/problems that punctuate their experiences. Our position is that "meaning" is a product of human beings actively experiencing the world as they "move" through time and space. As inherently social creatures, we attempt to share meaning linguistically (broadly cast) and generate artifacts through various technologies either as part of our communicating/collaborating with others or as an attempt to share meaning with another in the future who is in a similar situation or has a similar problem. The artifacts are what we call information. We believe that a useful way (if not the most useful way) to organize large numbers of such artifacts lies in the association between the active human being, his/her problem/situation and goal or envisioned end state and a co-collaborator (either in the present or in an envisioned future).

In Conclusion

The Internet, and more specifically the Web, has changed the functional definition of community by illustrating the collapse of the geographical/spatial constraint on communicating. Community implies an organizing principle more intimately tied to the problem/situation than to a person (or a set of people) per se, which has resulted in a multitude of virtual communities that are more narrowly focused on the range of problems/situations that define them. Along the way, the Web has also spawned a multitude of types of artifacts (what our field calls “information”) that come in a seemingly increasing number of unpredictable forms. Our inherited logic (so far at least) seemingly lacks the necessary coherency for the managing of these artifacts.

Another way to describe the Internet and its multi-media offspring, the Web, is the removal of many hierarchical or top-down constraints, both in terms of the “means of production” (à la Marx) and in terms of centralized organizational/political control (the Web would seem to be inherently un-manageable). This has resulted in an immense and rapid increase in the total volume of various types of artifacts. This will continue as various means by which people collaborate freely develop and evolve. This makes accurately predicting new ways of collaborating and what is created in terms of resultant artifacts very difficult (we would argue, impossible). However, the constant of fundamental human behaviors will remain. They will be manifested in different technological ways, yet potentially be more exposed than ever before. The i-field is seemingly the only community seeking profound insight into these behaviors through the lens that technology offers—albeit not formally so. We argue in this paper that anchoring ourselves in a constantly shifting bed of technology and artifacts will not lead to stability or necessarily profound insight into either of these, yet the proliferation of ways to communicate and resultant artifacts continues. This is why we believe a more focused and solidified agenda for the i-field based on fundamental human behaviors is not only timely, but—most importantly—necessary.

References

- Bourdieu, P. (1999). The social conditions of the international circulation of ideas. Richard Shusterman (Ed), Bourdieu; a critical reade, Oxford: Blackwell, 220-228.
- Bush, V. (1945). As we may think. The Atlantic Monthly.
- Carter, R. F. (1980). Discontinuity and communication, paper presented at the East-West Institute on communication theory from Eastern and Western perspectives, Honolulu, HI.
- Cohen, M. (2007). Insight from a delightful lunchtime conversation with Michael Cohen at the International Federation of Information Processing conference on Virtuality, Portland, OR, August.
- D'Eredita, M. A., Barreto, C. (2006). How does tacit knowledge proliferate? An episode-based perspective, *Organization Studies*, 27, 1821 –1841.
- D'Eredita, M. A., Nilan, M. S. (2007). Conceptualizing virtual collaborative work: Towards an empirical framework. K. Crowston, S. Sieber, E. Winn (Eds.), *International Federation of Information Processing*, Vol. 236, *Virtuality and Virtualization*, Boston: Springer, 21-34.
- Dervin, B. (1983). An overview of Sense-Making research: Concepts, methods and results to date. International Communication Association annual meeting, Dallas, Texas, USA.
- Dervin, B. (2003). Communication Gaps and Inequities: Moving Toward a Reconceptualization. B. Dervin & L. Foreman-Wernet (Eds), *Sense-Making Methodology Reader: Selected Writings of Brenda Dervin*, Cresskil, NJ: Hampton Press, Inc., 17-46.
- Dervin, B., Nilan, M. S. (1986). Information needs and uses, *Annual Review of Information Science and Technology*, 21, 3-33.
- Giddens, A. (1991). *Action, Structure, and Contradiction in Social Analysis*. Cambridge University Press.
- H.S. Kim, H. S. (2003). A theoretical explanation of collective life: Coorienting and communicating. B. Dervin, S.H. Chaffee and L. Foreman-Wernet, (Eds), *Communication: A Different Kind of Horse Race – Essays Honoring Richard F. Carter*, Cresskill, NJ: Hampton Press, Inc., 117-134.
- Jakobson, R. (1963). *Essais de Linguistique Generale: Volume 1 Foundations du Language*, Paris: Editions de Minuit.

- Logan, G. D. (1988). Toward an episode theory of automatization. *Psychological Review*, 95: 492-527.
- Logan, G. D. (1992) Shapes of reaction-time distributions and shapes of learning curves: A test of the episode theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18: 883-914.
- Nilan, M. S. (1992). Cognitive space: Using virtual reality for large information resource management problems. *Journal of Communication*, 42:4, 115-135.
- Nilan, M. S., D'Eredita, M. A. (2005). Organizations as virtual communities: A sense-making approach for uniting knowledge consumers and knowledge workers, *International Journal of Web Based Communities*, 1:3, 262-271.
- Nilan, M.S., Mundkur, A., Structuring virtuality: A user-based example for e-commerce. K. Crowston, S. Sieber, E. Winn (Eds), *International Federation of Information Processing*, Vol. 236, *Virtuality and Virtualization*, Boston: Springer, 53-66.
- Nonaka, I., Takeuchi, H. (1994). A Dynamic Theory Of Organizational Knowledge Creation. *Organization Science*, 5:1, 14 - 37.
- Reder, L. M. (1996) *Implicit Memory and Metacognition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Saracevic, T. (1997). Users lost: Reflections on the past, future, and limits of information science. *ACM SIGIR Forum*, 31:2, 16-27.
- Saracevic, T. (1999). Information science. *Journal of Information Science & Technology*, 50:12, 1051-1063.
- Sawyer, S., Huang, H. (2007). Conceptualizing information, technology, and people: Comparing information science and information systems literatures. *Journal of Information Science & Technology*, 58:10, 1436-1447.
- Taylor, R. S. (1986). *Value Added Processes in Information Systems*. Noorwood, NJ: Ablex.
- Weick, K. 1995. *Sensemaking in Organizations*. Thousand Oaks, CA: Sage Publications, Inc.
- Weick, K., & Roberts, K. 1993. Collective mind in organizations: Heedful interaction on flight decks. *Administrative Science Quarterly*, 38:3, 357-381.
- Wenger, E. (1998). *Communities of Practice*. New York: Cambridge University Press.
- Weisberg, R. W. (1993) *Creativity: Beyond the Myth of Genius*. New York, New York: W. H. Freeman and Company.
- Wyer, R. S. (1997) *The Automaticity of Everyday Life*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Yoon, K., Nilan, M.S. (1999). Toward a reconceptualization of information seeking research: Focus on the exchange of meaning. *Information Processing & Management*, 35, 871-890.

FULL PAPER SUBMISSION 22:

Reconsidering the practical application of facets on the web

AUTHOR(S):

La Barre, Kathryn

ORGANIZATION(S):

University of Illinois at Urbana-Champaign, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

BACKGROUND:

The term facet has become an increasingly common part of discourse about resource description and organization systems which are a critical part of contemporary information infrastructure. Facets seem to be intuitively adaptable structures for the practitioner who may integrate this concept into the practice of Information Architecture or the design of a Knowledge Management system (La Barre, 2006). As a browse and search device, facets seem to function equally well on e-commerce sites and in online library catalogs that use content management and search systems provided by vendors such as Endeca. Faceted access to folksonomies has been proposed as a way to tame unwieldy word clouds (Quintarelli, Resmini, Rosati, 2007). Facets and facet analysis are mentioned as valuable aspects of two critical discovery and access standards, NISO Z39.19, Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies (2005, pp. 14-15) and the current draft of Functional Requirements and Numbering of Authority Records (FRANAR) (2007, p. 23).

OVERVIEW:

Often absent from the current discourse are references to the theoretical structures that underpin the creation and use of facets. This omission obscures the fact that facets, as an information retrieval construct, arise from practice, from observation, and from use. The roots of facet theory lie in the creation of early information retrieval systems. One of the earliest faceted systems was created by the documentalists Otlet and La Fontaine in order to provide practical access to 'facts' in the form of bits of information pasted onto cards in a card catalog, with the hope of creating an immense synoptic outline of the world's knowledge (Rayward, 1997). Later, Ranganathan tested facets as an alternative to the hierarchical straightjacket of the DDC, by implementing a faceted access system in an academic library. After reclassing the library with his newly formulated colon classification, a faceted access system, Ranganathan and his staff observed patrons, tallied requests, and corrected issues long before fully formulating the principles of facet analysis. In Ranganathan's words (1971), "After a long experience is gained with an improvised aid, a theory is developed in order to understand the aid deeply and to systematize, improve, refine and develop it."

Given the heritage of facet theory, difficulties and issues that arise from current practice can be used to extend or strengthen the principles. Facets don't exist in a vacuum. Facet analysis must be anchored in anticipated use and in the applicable subject domain. North American understanding of facets is such that this device is often viewed as primarily useful for objects (not concepts) in small use domains. European understanding of facets is more nuanced, owing to greater experience with facets, dating from the early 1950s and their use, by subject specialists for organizing and creating access to special collections in business, government and special collections (Broughton, 2006).

METHOD:

This paper argues that clearly articulated principles for facets and facet analysis must be anchored in examples drawn from current practice in order to scaffold the development of fully developed faceted information infrastructures. Frequently, the term 'facet' seems to be used as a new way to describe pre-existing data fields in legacy systems such as bibliographic records and the backend relational databases on many e-commerce sites. In such a case it is not often clear that facet analysis, the critical first step in the construction of faceted access, has been conducted. Allowing the underlying data structures to fully determine all facets in use in a faceted navigation system may serve to inhibit the implementation of full featured systems. Other less obvious facets, or those that are not hardwired into an existing system, may emerge from conducting rigorous facet analysis and thereby permit iterative development of more powerful and dynamic systems.

This study will revisit a stratified random sample of 200 websites that were part of an earlier examination of the use of facets on websites. Changes in facet use will be noted using a content analytical protocol that was tested during the previous study (La Barre 2006). Websites from this sample that are found to use facets will also be compared with a set of 25, newly implemented 'faceted' online library catalogs. The author will conduct facet analysis on the content of each site in this subset following established facet analysis protocol (Aitchison, Gilchrist and Bawden, 2002; Atherton, 1961; Broughton, 2006). Differences between facets in use and facets that emerge from the process of facet analysis will be described. This process of comparison and analysis will be used to augment earlier evaluative guidelines that were previously extracted from failure analysis protocols used in testing earlier faceted systems (Atherton and Borko, 1965; Cleverdon and Keen, 1966; La Barre 2006).

DISCUSSION:

It can be difficult for an information architect or a knowledge management specialist to convince clients that time invested in detailed facet analysis is worth the cost (La Barre, 2006). This has resulted in partial implementations of faceted systems which are

now often used as exemplars to promote adoption of new faceted information access and discovery systems outside the business world. These existing systems have limitations that may be similar to the limitations of the systems they seek to replace, and may evince only superficial engagements with facet theory because they were developed in the time-bounded business culture that necessitates rapid return on investment (ROI). In this climate, without intervention, widescale adoption of faceted systems such as we are now seeing in online library catalogs may be fatally flawed from the outset. This paper will re-examine the extensive failure analysis and use studies, and principled application of facet analysis that were features of early information retrieval experimentation with faceted systems for evaluative guidelines useful to those considering faceted access systems. Rapid ROI is the order of the day in the economic climate that faces libraries and businesses. Will this prove to be the acid test of facets?

Many practitioners have requested that the residents of the academy fill these gaps with conscientious use and user testing in order to provide support for the anecdotal understandings by practitioners of the power of facets and facet analysis. Concerted, well designed tests like this are often impossible for in the business world as clients and upper management are too often unwilling, or unable, to pay for them. This intriguing symbiosis, between theory and practice provides a vital agenda item for I-school research during this period of rapid deployment and testing. In addition to the articulation of evaluative strategies for system implementations, this paper will outline a future agenda for formal and theoretically grounded comparisons of facet analysis with seemingly cognate traditions such as formal concept analysis and domain analysis. Information Science is engaged in a period of intensive experimentation and testing of the resource access and discovery mechanisms that will scaffold the information infrastructure of the twenty-first century. This paper outlines a research agenda that will draw from facet heritage and contemporary practice in order to ensure that facets are understood as practical extensions of theory, not dry, sterile, and devoid of connection to practice.

Aitchison, J., Gilchrist, A, and Bawden, E. (2002). *Thesaurus construction and use: A practical manual*. London: Fitzroy Dearborn.

Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings* 58 (1/2): pp. 49-72.

Atherton, P. (1961). A project for the development of a reference retrieval system for physicists (New York: American Institute of Physics, 1961) AIP/DRP 62-1.

Atherton, P., & Borko, H. (1965). A test of the factor-analytically derived automated classification method applied to descriptions of work and search requests of nuclear physicists. Report no. AIP/DRP 65-1; SDC/SP 1905. New York: American Institute of Physics.

Cleverdon, C. and Keen, M. (1966). *Aslib Cranfield research project - Factors determining the performance of indexing systems; Volume 2, Test results*. (Cranfield, EN: Aslib/ Cranfield Research Project, 1966) 173 also <http://hdl.handle.net/1826/863>.

La Barre, K. (2006). *The use of Faceted Analytico Synthetic Theory in website design and construction* (Unpublished dissertation). Indiana University, 2006

Quintarelli, E., Resmini, A; Rosati, L. (2007). Facetag: Integrating bottom-up and top-down classification in a social tagging system. *Proceedings 2007 IA Summit*. <http://www.facetag.org/download/facetag-20070325.pdf>

Ranganathan, S.R. (1971). *Colon Classification* 7th ed. Madras: Madras Library Association.

Rayward, W.B. (1997). The origins of information science and the International Institute of Bibliography/International Federation for Information and Documentation (FID) *JASIS* 48:4 pp. 289-300.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

Reconsidering the Practical Application

of Facets on the Web

Kathryn La Barre

University of Illinois at Urbana Champaign

501 E. Daniel Street, MC-493, Champaign,

IL 61820-6211 USA

1-217-244-4449

klabarre@uiuc.edu

ABSTRACT

The term facet has become an increasingly common part of the discourse about useful components of contemporary

information infrastructure. The facet approach to browsing or navigation as a way to scaffold resource discovery and access seems to be intuitively accessible to practitioners who

may integrate this concept into the practice of Information Architecture or the design of a Knowledge Management system (La Barre, 2006). Facets seem to function equally well on e-commerce sites and in online library catalogs using content browse, search or management systems provided by vendors such as Endeca or open source solutions such as Koha or solr. This paper discusses preliminary results of a survey of a stratified random sample of 200 websites that were part of an earlier examination in 2005 of the use of facets on websites. Changes in facet use over a two and a half year period were noted using a content analytical protocol tested during the previous study (La Barre 2006). Websites with evidence of facet use are compared with a set of newly implemented 'faceted' online library catalogs. Future research directions are also discussed.

Topics

- Information infrastructure development
- Information management
- Information organization
- Information technology and services

Keywords

Facet analysis, Website design, Information Retrieval, Content Analysis, Search and Discovery Systems, OPAC design.

50. INTRODUCTION

Facets are a fundamental part of a legacy information organization and access system known as faceted analytico-synthetic theory (FAST) that consists of a set of principles (facet analysis or FA), and techniques (faceted classification or FC) (Broughton, 2006). Together these principles and techniques can scaffold systems designed to organize and provide access to information by supporting search and discovery. A facet is "a generic term used to denote any component of a compound subject, also its ranked forms, terms and numbers" (Ranganathan, 1971, p. 88).

Contemporary use of the term facet extends this definition to include characteristics, features, aspects, or roles of a given information-bearing entity.

Facets seem to be intuitively adaptable structures for the practitioner who integrates this concept into the practice of Information Architecture or the design of a Knowledge Management system (La Barre, 2006). As recent implementations in library OPACs seem to indicate, facets function equally well as browsing or navigation devices on e-commerce sites and in library catalogs that use content management and search systems provided by vendors such as

Endeca or by open source software.

Support for faceted approaches appears in various forms. Faceted access to folksonomies has been proposed as a way to tame unwieldy word clouds (Quintarelli, Resmini, Rosati, 2007). Facets and facet analysis are also mentioned as valuable aspects of two critical discovery and access standards, NISO Z39.19, Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies (2005, pp. 14-15) and the current draft of Functional Requirements and Numbering of Authority Records (FRANAR) (2007, p. 23).

51. OVERVIEW

Often absent from current discourse are references to the principles and theoretical structures that underpin the creation and use of facets. This omission obscures the fact that facets, as an information retrieval construct, arise primarily from practice, from observation, and from use. The roots of facet theory lie in the creation of early information retrieval systems. One of the earliest faceted systems was created by the documentalists Paul Otlet and Henri La Fontaine in order to provide practical access to 'facts' in the form of bits of information pasted onto cards in a card catalog, with the hope of creating an immense synoptic outline of the world's knowledge (Rayward, 1997). Later, Ranganathan tested facets as an alternative to the hierarchical straightjacket of the DDC, by implementing a faceted access system in an academic library. After reclassing the Madras University library with his newly formulated colon classification (itself an early faceted search and discovery system), Ranganathan and his staff observed patrons, tallied requests, and corrected issues long before fully codifying the principles of facet analysis. In Ranganathan's words (1971), "After a long experience is gained with an improvised aid, a theory is developed in order to understand the aid deeply and to systematize, improve, refine and develop it."

Given the heritage of facet theory, difficulties and issues that arise from current practice can be used to extend or strengthen the principles and theory. Without such knowledge it can be easy to overlook the fact that facets don't exist in a vacuum. Facet analysis must be anchored in anticipated use and in the domain of application. North American understanding of facets is such that this device is often viewed as primarily useful for providing access to objects (not concepts) in small use domains which until recently were primarily commercial as earlier documentation retrieval experiments with facets were forgotten (Atherton, 1961; Atherton and Borko, 1965). Outside North America, understanding and use of facets is more nuanced, owing to greater experience and use of facets from the 1950s onward by subject specialists who organized and provided access to special collections in business, government and special libraries (Broughton, 2006).

The North American situation began to change drastically in early 2006 when North Carolina State University (NCSU) became the first library to implement the faceted or 'guided

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse this work. The authors are solely responsible for their

navigation' system of Endeca software in their library catalog (OPAC). Many other libraries followed suit, with the assistance of other vendors and the open source community. This heralds a re-entry into a time of exploration, exploitation and testing of the power of facets and various other approaches to enhancing the search and discovery experiences of those who use e-commerce websites and online catalogs. Many of the library catalog products are in beta versions, many are open source. A large number of new implementations are planned during the latter half of 2008. The NCSU implementation itself is not static. Functionality is still being added, but it remains one of the most complete demonstrations of approaches to faceted access currently available.

52. METHOD

Frequently, the term 'facet' is used as a new way to describe pre-existing data fields in legacy systems such as bibliographic records or the backend relational databases on many e-commerce sites. In such a case it is not often clear that facet analysis, the critical first step in the construction of faceted access, has been conducted. Allowing the underlying data structures to fully determine all facets in use in a faceted navigation system may serve to inhibit the implementation of full featured systems, especially in cases where the underlying data model does not fully exploit rich metadata, as is currently the case with bibliographic records. In such a situation, other, less obvious facets or those that are not hardwired into an existing system may emerge from conducting rigorous facet analysis. Such an approach permits iterative development of powerful and dynamic systems. In order to analyze new implementations of the faceted approach and to suggest improvements, clearly articulated principles for facets and facet analysis are necessary. It is helpful for examples of facet implementation strategies to be drawn from current practice in order to scaffold the development of fully developed faceted information infrastructures. For this reason, this research revisits a stratified random sample of 200 websites that were first examined in May 2005 as part of an earlier examination of the use of facets on the Web.

Observable changes in facet use on these sites were noted using a content analytical protocol that was previously tested (La Barre 2006). Websites with evidence of facet use were compared with a set of six newly implemented 'faceted' online library catalogs. Differences between facets in use on each set of sites are described.

Future work will include comprehensive facet analysis on the content of each 'faceted' site in this subset following established facet analysis protocol (Aitchison, Gilchrist and Bawden, 2002; Atherton, 1961; Broughton, 2006). Facets in actual use and facets that emerge from the process of facet analysis will be compared. This process of comparison and analysis will also be used to augment earlier evaluative guidelines that were previously extracted from failure analysis

protocols used in testing earlier faceted systems (Atherton and Borko, 1965; Cleverdon and Keen, 1966; La Barre 2006).

Sampling

The 200 sample websites in this exploratory survey were drawn from the Open Directory Project's (ODP) DMOZ directory (dmoz.org). At the time of the first survey in 2005 DMOZ consisted of seventeen main categories and contained links to over four million sites. It provided the core website data for Google and a number of other Web directories. DMOZ websites are chosen for inclusion by volunteer editors according to strict editorial guidelines. "ODP's goal is two-fold: to create the most comprehensive and definitive directory of the Web, and to create a high quality, content rich resource that the general public considers useful and indispensable." This seemed to offer a set of current websites that could provide a snapshot of common design practices. During the website survey phase of data collection in 2005, 200 sites were randomly selected from four DMOZ categories: (for-profit): Shopping, Business and (not for profit) Society and Reference in order to maximize comparability. Each randomly selected site was subject to a series of validation procedures to ensure currency, and comparability. Table 1 shows the figures for validation and stratified sampling of this group.

Table 1. Validation and stratified sampling for DMOZ sites

Site category	Top level sites	Validated sites	2005 Stratified sample
Reference (66,887)	n=544	n=106	13.5% (n=28)
Society (275,687)	n=305	n=196	25.1% (n=50)
Shopping (120,785)	n=543	n=224	28.7% (n=57)
Business (251,352)	n=719	n=255	32.7% (n=65)
Total (539,979)	(n=2111)	(n=781)	n=200

For the 2008 survey, websites from the 2005 survey were revisited. Those which remained online (n=170) and non-subscription formed the base set for comparison a set of six online library catalogs that self-identified as 'faceted catalogs'. Three are implementations of commercial vendor software and three are implementations of open source software as part of an effort to reflect the profit versus non-profit split evident in the DMOZ sample websites. Data for each survey (in 2005 and 2008) was collected over a two day period to reduce variability of results. The 2008 survey included a total of 176 sites.

Table 2. Sample of 'faceted' OPAC implementations (O/S denotes open source)

Endeca	North Carolina State University
Exlibris Primo	Vanderbilt University
OCLC Worldcat	U. Washington

Fac-Bac-OPAC (O/S)	Paul Smith's College
LibLime Koha Zoom (O/S)	Nelsonville Public Library
UVA Blacklight (O/S)	University of Virginia

Content analysis

Each of the sample sites from the 2005 and 2008 surveys were examined for a set of components, including search and browse areas, sitemaps, and navigational structures. A wireframe was created for each site; the basic and/or advanced search mechanisms were tested using the query 'search', the sitemap and first three levels of the website (including any search result screens resulting were printed out for later reference. Each wireframe was coded for the features described above and facets were identified.

The 2005 study found that designers often seem to be following Rosenfeld and Morville's (2002, p. 170) recommendation to integrate search and navigation (or browsing) systems on websites in order to allow users to jump easily between the two in order to conduct searches which lead to browsing, or searches which retrieve categories and documents. Thirty-six percent of the sites with integrated search and browse structures (see Figure 1 and 2) also used facets. "Instead of forcing one way to view the items, Faceted Navigation allows users to view the items in any way they want. At the same time they learn how the items are structured so that they may consider other search strategies in the future" (van Welie, 2004). Integrated search was also observed on each of the six library catalogs included in the 2008 study.

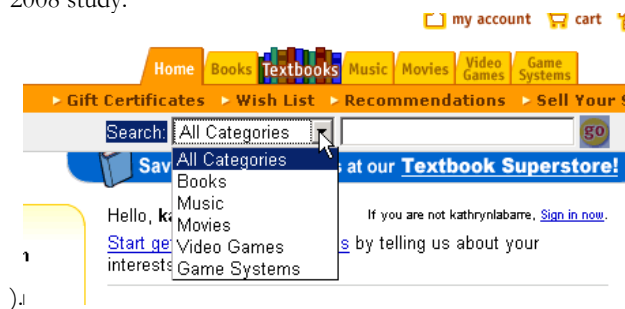


Figure 1. Integrated search/ browse feature on Amazon.com with type facet (books, music, etc.)

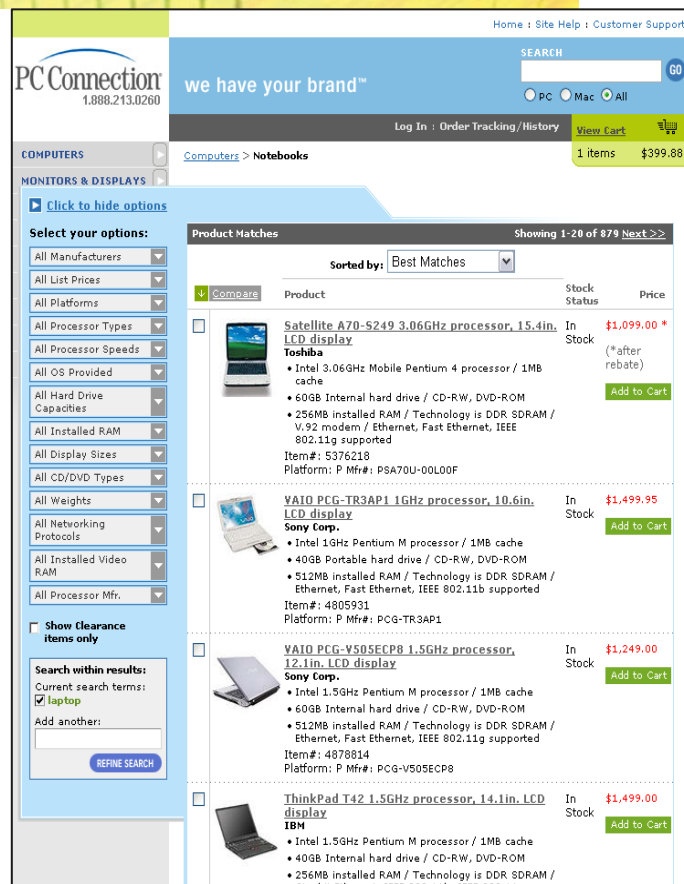


Figure 2. Integrated search/ browse feature using Endeca's 'guided navigation' on PC Connection.com with facets as options for browsing to narrowed result sets.

Uncovering facet use

Most (n=170) of the 176 sample websites are designed to provide access to concrete items, such as consumer products rather than subjects or ideas. Facets on all of the sites were uncovered by referring to Vickery's steps (1960, 1966) in facet analysis (1) Define the subject field by determining what entities are useful to the user (2) Formulate facets by examining material that expresses the needs of the user group to obtain candidate terms (3) Structure each facet by creating a hierarchical order and to identify gaps and overlapping concepts.

The U.S. Department of Labor page in Figure 3 is presented in order to demonstrate the coding and analysis protocol, and as an example of facet use on websites. Figure 3 is a screen capture of the entry page to the website and has a link to [Search/A-Z index]. This is page level 1 of a very complex website. Facets are in evidence as part of the navigation feature 'Find It'. A site user is invited to navigate the website by selecting an area of interest: topic, audience, popular items, form and location.

By following the page flow arrow (below) to the second level

results – as seen in Figure 5) Arrow 3 represents a page flow by use of the 'Find it' navigation structure.

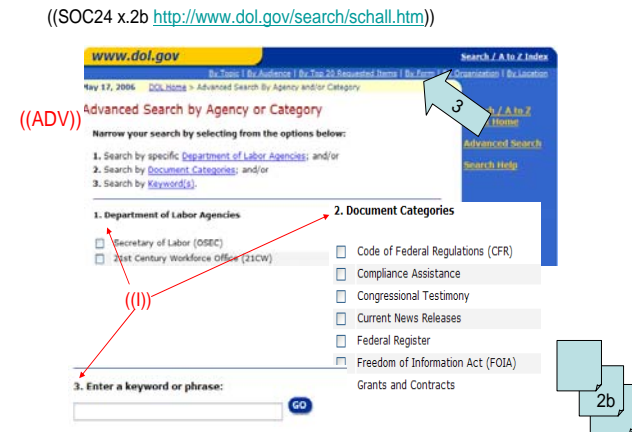
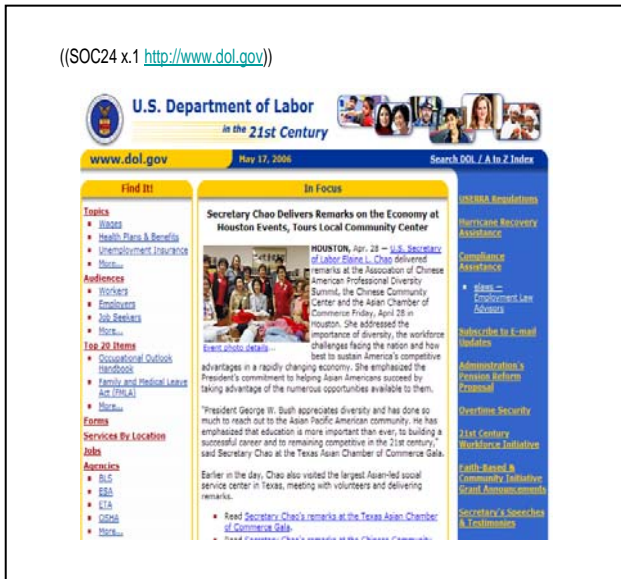


Figure 5. Advanced (integrated search) with facets

(2) of the site, two search components, basic and advanced are evident.

Figure 3. Department of Labor homepage.

Figure 4 is a wireframe that was created of the above site with facets and search features coded. It indicates that there are three page levels that were examined and each arrow indicates the direction of page flow.

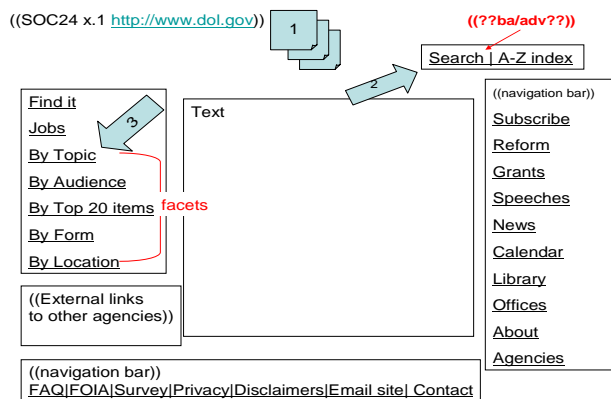


Figure 4. Wireframe of the Department of Labor homepage with facets indicated.

By following the page flow arrow in Figure 4 (below) to the second level (arrow 2) of the site, two search components, basic and advanced are evident. Because there are two kinds of search, basic and advanced, Arrow 2 can take the site user to either 2a (basic search results) or 2b (advanced search

Figure 5 demonstrates the way in which level 2b (advanced search) leads to an integrated search and navigation interface that allows users to refine or constrain a keyword search by agency and/or document categories. Document categories are not a good example of a facet (even though they are labeled as “categories”) because they are a jumble of document types and titles. This webpage does demonstrate the use of properly constructed facet attributes for Form (title, number and agency).

Were this use of faceting to be improved or more fully implemented, a display of types such as documents by code, regulation, origination, etc. would nudge this into more complete implementation. Facets are also drawn into use throughout the site on the 'Find it' navigation bar, in order to assist navigation.

Figure 6 (below) is a wireframe of the third level of the website as created from the screenshot from Figure 5. This figure also serves to demonstrate use of 'Form' facet attributes, by title, number and agency.

((SOC24 x.3 <http://www.dol.gov/libraryforms/>))

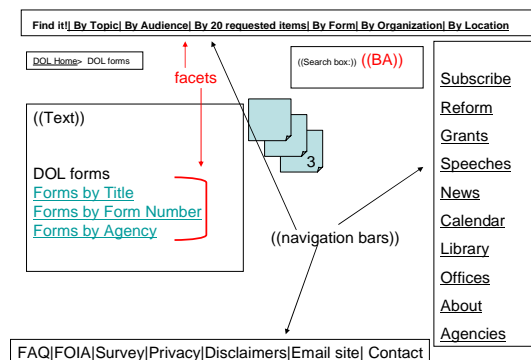


Figure 6. Page display when “Find it!” link [By Form] is followed.

53. Changes in the 2005 sample

Surprisingly, only 30, or 15% of the websites in the 2005 sample were offline, or had changed to subscription sites. This resulted in the loss of six previously identified faceted sites. Five of the offline sites were faceted. Another previously identified faceted website dropped from the sample when one of the shopping sites abandoned the use of facets in a redesign and removal of the browsing and

	2005		2008	
Reference	28	13.50%	23	13%
Business	65	25.10%	57	32.3%
Shopping	57	28.70%	51	28.9%
Society	50	32.70%	39	22.2%
OPAC	0		6	3.4%
	200		176	

advanced search features. Table 3 shows that over half of the remaining 170 sites had been updated in some way, and most had changed their navigation or search features, adding facets in the process. Business sites showed the greatest levels of updates to the sitemap, search and navigation components, while the Shopping sites showed the least number of changes to these components.

Table 3. Website changes (n=170)

	offline	updates
Reference	18%	56.50%
Business	11%	75.90%
Shopping	12%	43.10%
Society	22%	61.50%

Table 4 is a comparison of the 2005 and 2008 samples in terms of composition by category and across the two surveytime.

Table 4. Sample compositions between 2005 and 2008

As seen in Table 5, the number of sites in the sample decreased, while the remaining sites showed an increase in the use of facets on sites that had not previously used them.

Table 5. Changes in facet use by website category 2005 - 2008

	2005		2008	
Reference	2 (28)	7.10%	5 (23)	21.7%
Business	6 (65)	9.20%	18 (57)	31.5%
Shopping	9 (57)	15.80%	20 (51)	39.20%
Society	5 (50)	10%	7 (39)	17.90%
Total	22 (200)	11%	50 (170)	29.4%

Sites with new facet use tended to implement them as features in either browsing or searching interfaces. All new facet use on Reference sites were in the context of browsing interfaces. All new facet use in Society sites was implemented as part of the search interfaces. On Business and Shopping sites, new facet use was equally split among search and browsing areas. No sites from the 2005 sample showed new facet implementations in both areas.

54. Facet use in OPAC sample

The library catalogs chosen for inclusion in this website survey have had ardent supporters and implementers who have vigorously promoted each catalog. Table 2 lists the six catalogs which were included in the 2008 survey. Each of these implementations demonstrates improvements over the traditional e-commerce approach to facets. From comparing each of the implementations alone, it would appear that either the metadata in bibliographic records is far more complex than that on the e-commerce sites, or e-commerce metadata is severely underutilized.

Important differences emerge between e-commerce applications and library implementations when location of facet use and numbers of facets in use on each site are compared. These findings can only be indicative, not fully informative because of the small sample size.

Facets on the six library sites are used in both search and browsing components. This is rarely the case in the other categories of sites, as only one of the eighteen Business sites, and three of the twenty Society sites use facets in both areas. Average numbers of facets in use are revealing and range from an average low of 1.6 facets for search refinements use on the five Reference sites, to an average high of 4.4 in the seven browsing areas of the Business sites. Library use of facets in search interfaces averaged 10 per site, and 5.8 per site in browsing areas.

Table 5. Average facets in use by component and category. n=sites with facets in use. Parentheses indicate number of sites using each component per category.

	Searching	Browsing
Reference (5)	(1) 9	(5) 1.6
Business (18)	(12) 3.5	(7) 4.4
Shopping (20)	(13) 3.48	(10) 3.4
Society (7)	(5) 3	(2) 3

Library (6)	(6) 10	(6) 5.8
-------------	--------	---------

Facets used were also highly individual though some were evident across type of site.

Table 6. Most commonly used facets in the 176 sample sites

topic/subject	28	title	7
category	21	class number	6
form(at)	19	date/year	6
location	17	genre	6
brand	13	library	6
language	11	content	5
author	10	keyword	4
price	10	what	4
type	9	who	4
country/region	7		

While some of these facets may seem to be applicable only in libraries, some such as brand or price clearly seem to belong to the world of e-commerce. Facets such as topic/subject, format, language, author, location, date and year are used across all site categories and components.

Many questions remain for future work in this area. Might it be possible for library implementations to more completely harness the richness of the data in bibliographic records, while learning lessons from e-commerce implementations? Will the use of integrated search and browsing components be as useful to library users as they seem to be for e-commerce site customers? Are the relatively low numbers of facets used in the six implementations reflective of interface or software limitations? Are there generic facets that are useful across domain that are not being harnessed in websites? Are some facets more useful to aid in searching than browsing? Might facet analysis suggest facets not currently being used, but useful in given domains?

55. Conclusions

It can be difficult for an information architect or a knowledge management specialist to convince clients that time invested in detailed facet analysis is worth the cost (La Barre, 2006). This has resulted in a number of partial implementations of faceted systems such as Epicurious or Wine.com that are now used as exemplars to promote adoption of new faceted information access and discovery systems outside the business world. Existing and commonly referenced faceted systems have limitations that may echo the limitations of the systems they seek to replace. Too often these may evince no engagement with facet theory, or engage the principles superficially because they were developed in the time-bounded business culture that necessitates rapid return on investment (ROI). In this climate, without knowledgeable intervention, widescale adoption of faceted systems such as we are now seeing in online library catalogs may be fatally flawed from the outset. A re-examination of the extensive failure analysis and use studies, as well as principled

application of facet analysis that were common features of early faceted information retrieval experimentations may well be useful to those considering or implementing faceted access systems. Rapid ROI is the order of the day in the economic climate that faces libraries and businesses. Will this prove to be the acid test of facets? Many practitioners have requested that the residents of the academy fill these gaps with conscientious use and user testing in order to provide support for their anecdotal understanding of the power of facets and facet analysis. Concerted, well designed tests like this are often impossible in the business world as clients and upper management are too often unwilling, or unable, to pay for them.

This intriguing symbiosis, between theory and practice provides a vital agenda item for I-school research during this period of rapid deployment and testing. In addition to the articulation of evaluative strategies for system implementations, the future research agenda for this study is to provide formal and theoretically grounded comparisons of facet analysis with seemingly cognate traditions such as formal concept analysis and domain analysis. Information Science is engaged in a period of intensive experimentation and testing of the resource access and discovery mechanisms that will scaffold the information infrastructure of the twenty-first century. This paper outlined a preliminary research agenda that draws from facet heritage and contemporary practice in order to ensure that facets are understood as practical extensions of theory, not dry, sterile, and devoid of connection to practice.

56. ACKNOWLEDGMENTS

These preliminary results were presented at the panel: Reading the Information Infrastructure: Functions, Facets, and Tags panel discussion that included two additional papers: The Social (Tagging) Act of Reading by Cheryl Knott Malone, panel organizer, and FRBR as an interdisciplinary high-middle range theory by Allen Renear and Dave Dubin.

57. REFERENCES

- [81] Aitchison, J., Gilchrist, A, and Bawden, E. (2002). Thesaurus construction and use: A practical manual. London: Fitzroy Dearborn.
- [82] Atherton, P. (1961). A project for the development of a reference retrieval system for physicists (New York: American Institute of Physics, 1961) AIP/DRP 62-1.
- [83] Atherton, P., & Borko, H. (1965). A test of the factor-analytically derived automated classification method applied to descriptions of work and search requests of nuclear physicists. Report no. AIP/DRP 65-1; SDC/SP 1905. New York: American Institute of Physics.
- [84] Beccaria, M., and Scott, D. 2007. Fac-Back-OPAC an open source interface to your library system. Computers in Libraries. 27(9) DOI=http://www.infotoday.com/cilmag/oct07/Beccaria_Scott.shtm

- [85] Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings* 58 (1/2): pp. 49-72.
- [86] Cleverdon, C. and Keen, M. (1966). *Aslib Cranfield research project - Factors determining the performance of indexing systems; Volume 2, Test results.* (Cranfield, EN: Aslib/ Cranfield Research Project, 1966) 173 also <http://hdl.handle.net/1826/863>.
- [87] La Barre, Kathryn (2006) The use of faceted analytico-synthetic theory as revealed in the practice of website construction and design. Ph.D. dissertation, Indiana University, United States -- Indiana. Retrieved 25-1-08, from ProQuest Digital Dissertations database. (Publication No. AAT 3219917).
- [88] Quintarelli, E., Resmini, A; Rosati, L. (2007). Facetag: Integrating bottom-up and top-down classification in a social tagging system. *Proceedings 2007 IA Summit*. <http://www.facetag.org/download/facetag-20070325.pdf>
- [89] Ranganathan, S.R. (1971). *Colon Classification* 7th ed. Madras: Madras Library Association.
- [90] Rayward, W.B. (1997). The origins of information science and the International Federation for Information and Documentation (FID) *JASIS* 48:4 pp. 289-300.
- [91] van Welie, M. (2004). Web design patterns: Faceted navigation. Retrieved: 2 February 2004 from <http://www.welie.com/patterns/showPattern.php?patternID=faceted-navigation>

FULL PAPER SUBMISSION 23:
Discipline Formation and the Field of Information

AUTHOR(S):
Miksa, Fran

ORGANIZATION(S):
University of Texas at Austin, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The field of information as expressed in the iSchool Caucus is in many respects a conglomeration of disparate elements. It includes, for example, elements of several major traditions of information organization (commonly expressed as service professions) and the social institutional ties and contexts that some of them entail. In addition the field has also gathered under its umbrella an impressive array of such other elements as aspects of cognitive studies, social studies of the users and use of information, studies and services related to the Internet and World Wide Web, digital libraries and other systems that have no particular social institutional setting, digitization initiatives, not least among which are growing museum and digital archives movements, studies related to information systems of all kinds (e.g., interface design, usability testing, security), social issues of information property and rights, and the entire array of informatics initiatives, to name only some of the whole. Within this enlarged context, one issue of significance is the role of discipline formation, where the latter refers to efforts to identify a fundamental phenomenon related to larger core professional activities and explain aspects of that phenomenon through scientific methodology and objectivity.

The purpose of this paper is to use two examples of discipline formation within traditions of information organization (Bibliography and Computerized information storage and retrieval) as a basis for enriching and provoking our contemporary understanding of discipline formation within the field of information. As such, it focuses on the fourth review criterion of the iSchool Caucus program: “develops intellectual geographies in which attendees can learn about intellectual domains not their own but part of the multi-disciplinary iSchool space.”

Background

A tradition consists of a cluster of activities, objects, and ideas that are handed down from generation to generation in human society. Each such tradition gains adherents and subsequently evolves and shifts through endogenous and exogenous factors. A tradition of information organization has to do more specifically with activities and ideas related to organizing information objects and the information embedded in such objects. Traditions of information organization go beyond practical techniques. For example, they include ideas related to values that are attributed by their adherents to information objects as objects and the information they contain and to the roles and uses of such objects and their informational content in society. All major traditions of information organization in the modern period have had their beginnings in the form of what today would be called service professions.

The Two Examples

Two traditions of information organization—Bibliography (over five centuries old) and Computerized information storage and retrieval (CISAR), the latter encompassing a wide variety of activities developed in the post-WWII era—provide a basis for raising questions about discipline formation within the field of information. Salient features include the fact that each took as its beginning point the identification, representation and item-by-item listing of documents deemed of value for their subject content. Each initially took on professional service ideals related to the crisis of providing access to burgeoning publications, especially in the realm of scholarly and scientific research. Despite such service profession orientations and unlike other traditions of information organization, each has been relatively free from the constraints of a specific social institutional setting. Each went through periods of attempting to describe and clarify what amounted to uncertain boundaries and often-changing content. Each developed commercial expressions of its work, though in the case of Bibliography commercial expressions came to be expressed chiefly in the form of two traditions of information organization—Indexing and Documentation—for which it was the principal seedbed.

Most significantly, each of these traditions gave rise to discipline formation within its boundaries—Bibliography in the form of Analytical (or Critical) bibliography, CISAR in the form of Bibliometrics and Computational information retrieval. Each effort in discipline formation had both positive and negative effects. Positive effects consisted of the satisfaction gained by its adherents of

being able to state more precisely what their realm of scholarly endeavor entailed, an increase in scholarly status within the general academic community, and attendant increases in public acceptance and funding that an improved status generated. Negative effects included various expressions of alienation of the adherents in the parent tradition when those forming a disciplinary focus in this way attempted to distance themselves from the parent tradition, and the effect of criticisms of the scientific base of the new discipline that were generated because of the conflict involved.

Why This Matters

Discipline formation within the field of information seems inevitable or at least highly probable (if not already well underway) given the status that disciplinary work generates in the modern world. And given that inevitability or high probability, it seems appropriate to try to apply at least some of what occurred in the two instances described here to the current information field situation. However, taking such notice of such past developments and applying the latter to the present situation can best be done not in the form of predictions which are all but impossible, but by attempting to pose useful questions based on or at least provoked by relevant issues in the past.

A list of more general topics around which such questions might revolve include the existence of professional service ideals and their relationship to the idea of an academic discipline focus, the problem of defining boundaries and content of the field as a whole or of some elements of the field (e.g., an information organization tradition within the field), and the role of commercial parameters of the field. Topics more specific to discipline formation itself might include the importance of identifying a fundamental phenomenon to which rigorous scientific methods and explanation may be applied, how differing phenomena might yield different discipline formation efforts and how they might co-exist, and both positive and negative effects of discipline formation for the field as a whole. More specific questions might, for example, take the form of 1) Can information service profession needs co-exist with discipline formation within the information field and, if so, how? 2) What are the costs of severing ties (for various reasons) with parent traditions of information organization and their social institutional settings? 3) What might be expected in the way of idea interplay regarding the scientific basis and objectivity of a discipline within the larger information arena? 4) What might be the stance of the information field in general towards discipline formation that downplays scientific methodologies and explanations? Suggestions will be made about how to think about answers to questions such as these, though the purpose of raising them as well as reciting the comparison provided here that lies behind them and that prompted them is to provoke discussions as to what questions regarding discipline formation within the information field are the most important.

FULL PAPER SUBMISSION 24:

Designing E-Learning Games on Cellphones to Promote Language Learning and Literacy in the Developing World

AUTHOR(S):

Kam, Matthew

ORGANIZATION(S):

University of California at Berkeley, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

PAPER ABSTRACT – iConference 2008

Designing E-Learning Games on Cellphones to Promote Language Learning and Literacy in the Developing World

Literacy is one of the great challenges in the developing world. Even more challenging is the tension between regional and “world” languages – that economic opportunities are often closed to those who are literate only in a regional language. India for instance is a country with 22 regional and two “national languages” Hindi and English. But for most practical purposes, English is the language of economic opportunity. It is the language of instruction in private schools and all universities, a large fraction of business and government, and the language which is driving India’s service economy – the fastest growing in the world. Fluency in English can almost be equated with membership in the middle and upper classes [6]. The value of English is widely recognized by ordinary Indians [15], and it is the poorest citizens who are lobbying most strongly to expand English teaching. A recent article states that mastery of English is the “single most influential factor that determines access to elite educational institutions, and hence to important avenues of economic and social advancement” [9].

More broadly, the literature [e.g. 6], our conversations with development professionals in Africa, Latin America, and Asia, and further experiences in the field indicate that a large proportion of low-income people want to improve their command of an appropriate “world language.” English is certainly one of these, as is Mandarin Chinese and Spanish. But even in countries where such a language is an official “national language,” many speakers (inevitably the least empowered) have a different native language, and many *regional* languages (let alone dialects) are often spoken. In India, Hindi and English are official “national languages,” but Hindi is native to only 20% of the population. “World language” fluency opens the door to further education, a larger regional (or world) marketplace, to “new economy” outsourced jobs, and often improves access to government, health and legal services.

Our current work concentrates on ESL (English as a Second Language), but we believe many of the lessons will transfer to other “WSL” (World language) learning challenges. Unfortunately, formal English teaching in public schools is not succeeding, and it is far out of reach to the large numbers of children who are not able to attend school regularly. From the literature [e.g. 1] and our fieldwork in the poorest state of India, two significant factors stand out: irregular school attendance owing to the need for students to work in fields, homes, etc., and the preparation of local teachers, who themselves have very poor training in English. Regional schools typically have no “English teacher,” rather a handful of teachers who cover everything. In our 3 years of fieldwork, we have usually been unable to communicate with local teachers in English and have relied on interpreters.

At the same time, high technology – in the form of cellphones – are being used by Indians at all economic levels. India is the largest market for cellphones worldwide, and the majority are now being bought by illiterate and semi-literate users. And a growing percentage of these phones feature advanced multimedia capabilities for photos and gaming. We argue that these devices are a perfect vehicle for new kinds of informal (out-of-school) language learning. All these factors create an extraordinary opportunity: to dramatically expand English skills in young Indians, which is the fastest way to open the doors for employment and further education. While our

PAPER ABSTRACT – iConference 2008

work can be integrated with, and can complement the school English curriculum, the greatest opportunity is in out-of-school learning. The Indian children we studied, even those with work commitments for the family, spent an hour or so daily playing with friends, and several hours watching television.

We believe that ESL learning games on cellphones can address the above challenges. Games have well-known “immersive” properties – that is, the learner experiences themselves as being “inside” the game. Games continually challenge players to develop new skills, and reward their acquisition. Players often spend long hours in games which allows them to develop advanced skills. Games can be social experiences which encourage communication with other players. And finally, games can be designed to mirror the real-world games that children play, providing them with tools and encouragement to use English recreationally *outside* the electronic game.

Leveraging e-learning games for education in developing regions is not far-fetched. At least two non-government organizations, Pratham and the Azim Premji Foundation, have used computer games in their initiatives for children in the urban slums and rural areas of India respectively. Most importantly, a large-scale evaluation by Pratham¹ showed significant gains on mathematics test scores from playing computer games that target math learning [2]. It is plausible that similar learning outcomes can be replicated using mobile games for ESL.

In this paper, I introduce my dissertation work, which is now in its 4th year. It is informed by six rounds of fieldwork over three years with children living in the urban slums and rural areas of India. Our total time actually in the field is over six months. A goal of these field studies was to learn first-hand about the everyday learning contexts of the above children as part of a broader needs assessment. Another objective was to examine the feasibility of our ideas for technology-assisted learning, by piloting off-the-shelf software and early designs of our mobile applications on a very small scale with children from the urban slums and villages. Building on the systems requirements that we have learned iteratively from these studies, we are completing a final round of design and prototyping in Fall 2007 in preparation for a year-long deployment in 2008.



Rural children displaying “You win!” screens from two cellphone games that we have designed. Success in these games, and mastery of the technology, was a considerable incentive and source of pride for these children. These kinds of behaviors were repeated over and over.

¹ A longitudinal randomized experiment over more than two years with over 10,000 urban slums students.

PAPER ABSTRACT – iConference 2008

Since our previous findings are already published, I will only review those games that we have designed so far and their learning outcomes. Instead, in this paper, I want to focus on meta-issues related to the project's design methodology and its interdisciplinary nature. In particular, I want to draw attention to the challenges of doing information systems research in this highly difficult domain and offer strategies based on how I would undertake a project in this domain differently if I had to do it a second time.

The rest of this paper addresses the following issues:

- What is an effective model for multidisciplinary collaboration? While the most conventional approach is to assemble a specialist from each domain to form a team, we have learned that specialists in their respective areas are not always able to relate to the unique problem that lies at the intersection of these multiple areas of expertise. I have often found myself acting as the “interpreter” between computer programmers and curriculum developer in my team. More importantly, I will describe the theoretical frameworks that I drew upon from second language acquisition [3, 4, 6, 10, 11, 12, 16] and game studies [8, 13, 14, 17], and how my knowledge of both areas (to a certain degree) in addition to human-centered design made it possible for me to perform the essential “conceptual footwork” for integrating them in our e-learning game designs, even as I relied on our curriculum developer for her practical experience in teaching English in India.
- How can we develop competencies across disciplinary boundaries that make interdisciplinary work possible? Here, I will trace my intellectual development from my initial schooling in computer science and human-computer interaction to my realization that the problem I am confronting in my dissertation actually belongs to the education domain, and my subsequent socialization in the School of Education to learn their “trade.”
- What are some appropriate approaches for conducting cross-cultural research in information systems? While ethnographic methods are critical for understanding the practices and beliefs of a community, culture is a dynamic phenomenon that changes quickly especially in the face of rapid technology uptake (e.g. cellphones) in emerging regions. I will cite my experiences from an international development project in Africa involving handheld technology to argue that conducting feasibility studies with multiple user communities is necessary for revealing our cultural assumptions. In particular, I will support this claim by describing the differences in how children in North and South India play our games differently.
- What data collection challenges are there with users in developing regions who lack the exposure to technology and literacy levels for them to articulate their thoughts meaningfully so as to be *directly* useful for technology design? We have found that while it is difficult to obtain useful data from child users other than by observing them use the technology, we have obtained useful feedback by showing our designs to highly educated, local informants who know these children well as caregivers or educators. Similarly, the mundane obstacles in conducting fieldwork with local communities with the help of interpreters offer clues to ingrained beliefs and deeper patterns of behavior that affect technology use. I will describe a few of these behaviors that we have observed.

PAPER ABSTRACT – iConference 2008

REFERENCES

1. Azim Premji Foundation. The Social Context of Elementary Education in Rural India, 2004.
<http://www.azimpremjifoundation.org/downloads/TheSocialContextofElementaryEducationinRuralIndia.pdf>.
2. Banerjee, A., Cole, S., Duflo, E., and Lindon, L. Remedying Education: Evidence from Two Randomized Experiments in India. *NBER Working Paper No. 11904*, December 2005.
3. Bialystok, E., and Hakuta, K. *In Other Words: The Science and Psychology of Second-Language Acquisition*. Basic Books, 1994.
4. Cameron, L. *Teaching Languages to Young Children*. Cambridge University Press, 2001.
5. Clegg, J., Ogan, B., and Rodseth, V. Evaluating Digital Learning Material for English Language Development in African Primary Classrooms. *IMFUNDOKnowledgeBank Paper*, April 2003.
6. Ellis, R. *Second Language Acquisition*. Oxford University Press, 1997.
7. Faust, D., and Nagar, R. Politics of Development in Postcolonial India: English-Medium Education and Social Fracturing. *Economic and Political Weekly*, India, July 28, 2001.
8. Gee, J.P. *What Video Games Have to Teach Us About Learning and Literacy*. Palgrave Macmillan, New York, USA, 2003.
9. Kishwar, M.P. *Deprivations's Real Language*, 2005.
<http://www.indianexpress.com/printerFriendly/12662.html>.
10. Larsen-Freeman, D. *Techniques and Principles in Language Teaching*, 2nd Edition. Oxford University Press, Oxford, UK, 2000.
11. Lightbown, P., and Spada, N. *How Languages Are Learned*. Oxford University Press, 2006.
12. Nunan, D. *Task-Based Language Teaching*. Cambridge University Press, 2004.
13. Rouse, R. *Game Design: Theory and Practice*. Wordware Publishing, 2001.
14. Salen, K., and Zimmerman, E. *Rules of Play: Game Design Fundamentals*. The MIT Press, 2004.
15. Shukla, S. From Pre-colonial to Post-Colonial: Educational Transitions in Southern Asia. *Economic Political Weekly*, 1996. 31(22), 1344-49.
16. Snowling, M.J., and Hulme, C. (Eds.) *The Science of Reading: A Handbook*. Blackwell Publishers, 2005.
17. Sweetser, P., and Wyeth, P. GameFlow: A Model for Evaluating Player Enjoyment in Games. *ACM Computers in Entertainment*, 3(3), July 2005.

FULL PAPER SUBMISSION 25:

An Exploration of the Life Cycle of eScience Collaboratory Data

AUTHOR(S):

Wallis, Jillian C
 Pepe, Alberto
 Borgman, Christine L
 Mayernik, Matthew S

ORGANIZATION(S):

University of California, Los Angeles, United States of America
 Center for Embedded Networked Sensing

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The success of eScience research depends not only upon effective collaboration between scientists and technologists but also upon the active involvement of data archivists. Archivists rarely receive scientific data until findings are published, by which time important information about their origins, context, and provenance may be lost. A better understanding of these processes will enable archivists to participate in earlier stages of the life cycle and to improve curation of these types of scientific data.

Research reported here addresses the life cycle of data from collaborative ecological research with embedded networked sensing technologies. Evidence from our interview study and field research yields a nine-stage life cycle. Among the findings are the cumulative effect of decisions made at each stage of the life cycle; the balance of decision-making between scientific and technology research partners; and the loss of certain types of data that may be essential to later interpretation.

Archivists typically receive scientific data only after the findings of a study are published or after a researcher retires. Neither of these archival outcomes provides access to scientific data in a timely manner. More importantly, by the time that archivists receive data, much of the information necessary for future interpretation may have been lost. Shifting the practices of archiving such as appraisal, curation, and tracking provenance into earlier stages of a given material's life cycle can increase the likelihood of capturing reliable, valid, and interpretable data and thus improve both short- and long-term access and interpretation.

To determine how early these archiving processes might begin, it is necessary to identify the life cycle of a given type of data. We are not alone in this pursuit -- in keeping with the scientific data archival research agenda for the next decade, our goals are to develop: more detailed data models for each domain, including intra-domain and inter-domain commonalities, automatic processes for data and metadata capture, and consistent methods of data description in this scientific and technical environment.

Summary of Life Cycle Stages

1. Experiment Design. The beginning of the data life cycle is the design of new experiments. CENS researchers design new experiments by reusing data from prior research.
2. Instrument Calibration & Ground-Truthing. Before sensors are deployed, they are calibrated to known solutions or values to identify the offset between the actual measurement and the expected measurement.
3. Data Capture. Once sensors have been deployed successfully in the field, researchers begin to collect observations of physical phenomena. Some sensor measurements are direct (e.g., temperature, wind speed) and others are indirect (e.g., measure of fluorescence as an indicator of chlorophyll activity).
4. Cleaning Data. After data have been captured, calibration and ground-truthing information need to be applied to the data to normalize any calibration offsets from the sensing equipment.
5. Data Integration. Few of the observations and samples collected in the field can be interpreted without derivation into more meaningful data points. Data typically must be averaged into composite points before they can be used in analysis.
6. Derivation. Researchers are looking for trends over time and across spatial locations. Datasets each given deployment are integrated by multiple researchers, for multiple reasons, and in multiple combinations.
7. Analysis. Researchers use statistical, modeling, and visualization tools that vary by research specialty and individual preference. They test and generate hypotheses and draw conclusions about data obtained from the deployments.
8. Publication. Data collected during embedded network sensor deployments culminate in scholarly publications such as journal articles, conference papers, posters, and technical reports.
9. Preservation. Few, if any, of the CENS researchers interviewed had data preservation strategies commensurate with those of the archival community. It is more accurate to say that they back up their data.

Summary of Conclusions

We have identified nine stages that appear to be common to the CENS deployments studied, the researchers, and to the resulting data. The order of the steps is not absolute, as some stages are iterative while others may occur in parallel. Actions taken at each

stage of the life cycle influence how the resulting data can be interpreted, hence it is important that these stages be documented and associated with the resulting dataset. There is a cumulative effect of decisions made at each stage of the life cycle. For example, decisions made in the experimental design stage determine what data exist for analysis, or calibration decisions are essential to interpreting the data. There is a balance of decision making between scientific and technology research partners. The engineering data may be more essential to later interpretation of application data than previously believed.

In our ongoing research we are exploring and refining the data life cycle identified here, as well as build systems to support it. For instance, we need to understand more fully the true interplay of the cleaning, integration and derivation stages. We are also curious about the role of decisions made in the field on the calibration and capture interplay, as well as the experiment and design interplay. At present, much of the sensing technology is experimental, but commercial off-the-shelf sensors also are in use. Research questions about data provenance will evolve as the technology stabilizes and the scientific research questions broaden.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

- practices in habitat ecology. in 10th European Conference on Digital Libraries. 2006. Alicante, Spain: Berlin: Springer.
- [12] Borgman, C.L., J.C. Wallis, and N. Enyedy. Little Science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, 2007.
- [13] Mayermik, M.S., J.C. Wallis, and C.L. Borgman. Adding Context to Content: The CENS Deployment Center. in *American Society for Information Science & Technology*. 2007. Milwaukee, WI: Information Today.
- [14] Borgman, C.L., et al. 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*. 2007. Vancouver, BC: Association for Computing Machinery.
- [15] Pepe, A., et al. *Knitting a fabric of sensor data and literature*. in *Information Processing in Sensor Networks*. 2007. Cambridge, MA: Association for Computing Machinery/IEEE.
- [16] Wallis, J.C., et al. Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research. in *3rd International Digital Curation Conference*. 2007. Washington, D.C.
- [17] Glaser, B.G. and A.L. Strauss, *The discovery of grounded theory: strategies for qualitative research*. Observations. 1967, Chicago: Aldine Pub. Co. x, 271.

An Exploration of the Life Cycle of eScience Collaboratory Data

Jillian C. Wallis	Alberto Pepe	Matthew S. Mayernik	Christine L. Borgman
Center for Embedded Networked Sensing, UCLA 00+1+3102060029 jwallisi@ucla.edu	Dept of Information Studies Graduate School of Education & Information Studies, UCLA 00+1+3102060029 apepe@ucla.edu	Dept of Information Studies Graduate School of Education & Information Studies, UCLA 00+1+3102060029 mattmayernik@ucla.edu	Dept of Information Studies Graduate School of Education & Information Studies, UCLA 00+1+3108256164 borgman@gseis.ucla.edu

ABSTRACT

The success of eScience research depends not only upon effective collaboration between scientists and technologists but also upon the active involvement of information scientists. Archivists rarely receive scientific data until findings are published, by which time important information about their origins, context, and provenance may be lost. Research reported here addresses the lifecycles of data from ecological research with embedded networked sensing technologies. A better understanding of these processes will enable information scientists to participate in earlier stages of the life cycle and to improve curation of these types of scientific data. Evidence from our interview study and field research yields a nine lifecycle phases, and three types of lifecycle depending on the research goal. Findings include highlighting the impact of collaboration on the research processes and potential phases during which the integrity of the captured data is compromised.

Topics

Cultural information systems, Information infrastructure development, Information management, Preserving digital information

1. INTRODUCTION

The success of eScience research depends upon effective collaboration between scientists and technologists. Partners often must learn how to produce data that are meaningful to participants from multiple disciplines. Many decisions are made about data at each stage in its life cycle. Curation of these data and their value for reuse depends heavily on how much is known about their origins, derivation, and provenance.

Archivists typically receive scientific data only after the findings of a study are published or after a researcher retires. Neither of these archival outcomes provides access to scientific data in a timely manner. More importantly, by the time that archivists receive data, much of the information necessary for future

interpretation may have been lost. Shifting the practices of archiving such as appraisal, curation, and tracking provenance into earlier stages of a given material's life cycle can increase the likelihood of capturing reliable, valid, and interpretable data [1] and thus improve both short- and long-term access and interpretation.

To determine how early these archiving processes might begin, it is necessary to identify the life cycle of a given type of data. eScience partners often have different responsibilities at each stage of a life cycle. Individual researchers may be insufficiently aware of how others have acted upon the data, or how others may use or interpret the data further down the line. Making the entire life cycle of data more transparent and self-documenting has the potential to simplify data capture, management, interpretation, and curation for all parties involved [2, 3]. Some stages can be augmented by technical means, such as automated tools to identify potential instrumentation errors as they occur. Other stages can be made more transparent by identifying and documenting scholarly practices associated with the data.

The life cycle of business and government documents is characterized by each stage being handled by a different party. The life cycle of data from little science – that is, science performed by an individual or small research group – is characterized by all of the phases being handled by one or a few persons with similar domain knowledge and training. The life cycle of data from big science – that is, science performed by a large number of researchers, such as high-energy physics – is characterized by many researchers participating in each stage of the life cycle. These researchers all have similar domain knowledge and training. In the research reported here, researchers from multiple disciplines play complementary (and sometimes conflicting) roles in data handling.

In keeping with the scientific data research agenda for the next decade set by the Warwick Workshop [4], our goals are to develop: a) more detailed data models for each domain, including intra-domain and inter-domain commonalities, b) automatic processes for data and metadata capture, and c) consistent methods of data description in this scientific and technical environment. Our exploration of the life cycle of scientific data identifies the stakes and stakeholders at each phase to develop a “digital curation infrastructure” [5] that will support the use, reuse, access, and interpretation of ecological sensing data. In this context, we need to understand the processes that lead to the creation, analysis, and publishing of said data for metadata capture, and when major changes occur to data so that we can

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse this work. The authors are solely responsible for their paper's content. Our thanks to the Association for Computing Machinery for permission to adapt and use their template for the iSchools 2008 Conference.

build appropriate provenance tracking measures. Born-digital objects leave no physical residue that can be referenced later, too often, useful information is discarded before being properly assessed for archival value [6].

2. BACKGROUND

Research reported here is affiliated with the Center for Embedded Networked Sensing (CENS), a National Science Foundation Science and Technology Center established in 2002 [<http://www.cens.ucla.edu/>]. CENS supports multi-disciplinary collaborations among faculty, students, and staff of five partner universities across disciplines ranging from computer science to biology. The Center's goals are to develop and implement wireless sensing systems, and to apply this technology to address questions in four scientific areas: habitat ecology, marine microbiology, environmental contaminant transport, and seismology. Application of this technology already has been shown to reveal patterns and phenomena that were not previously observable.

Our data management research group has been part of CENS since its inception. While few scientific data were generated in the early years, we were planting the seeds of archival practice and preservation. Once data captured by CENS' instrumentation became relevant to our application scientists, we took a more active role in building the necessary infrastructure for long-term access. Our initial research focused on defining what were "data" in this environment. Now that we understand better what are data to whom and when, we are addressing larger data life cycle issues.

2.1 Deployment Scenario

An example of a CENS embedded networked sensing system deployment will provide context for the life cycle of CENS data.

CENS researchers utilize several deployment models. Along with static deployments typical of observatories such as NEON or GEON [7, 8], CENS researchers regularly go on short-term deployments, or "campaigns," where sensing systems are deployed in the field for a few days. Among the benefits of this approach for exploratory research are compatibility with the data collection practices of application science researchers (most are in biology or environmental sciences), the ability to field test delicate and expensive experimental equipment, and the opportunity for science and engineering researchers to work together in the field to trouble-shoot technical problems and improve the overall quality of data.

An example of a CENS deployment is the study of biological processes associated with harmful algal blooms. In designing a deployment, the application science researchers (biologists in this example) identify a viable research site, in this case a lake known for summer blooms. Available background information about the lake includes peak months for algae, a topology of the lakebed, local species of phyto and zooplankton, and nutrient presence and concentration. The engineering researchers determine which equipment are most appropriate for capturing the data desired by the scientists.

Prior to going in the field, the team calibrates equipment in the laboratory based on knowledge of the types of organisms likely to be present in the water. Because of the natural variation of water organisms, calibrations will be augmented with physical water samples taken adjacent to sensors. A "wet lab" will be set up on site to process water samples. Once on site, the team deploys

sensors in the lake using static buoys that house a power source, data logger, and wireless communication system. They document GPS coordinates of each buoy, times of placement, and serial numbers of each sensor in a laboratory notebook.

The data collection process is a combination of pre-planned activities and in-field decisions. Because the aquatic phenomena of interest vary on diel or 24-hour cycle, scientists take data for a full 24 hours. Once sensors begin to report data, researchers begin observing interesting phenomena, such as that the water flows more quickly at one end of the lake, and that the water is greener and at a higher temperature where a rock slows the flow. Based on such information, the team may change the data collection strategy, altering plans for sensor placement or for hand collection of water samples. At the end of a deployment, equipment is removed and returned to the lab. Water samples are processed for organism identification and concentration and for nutrient concentrations. Sensor data are compared to the in-lab and in-field calibration curves and to other trusted data sources. Only then are water sample data and sensor data integrated for analysis. After data analysis is complete and papers are published, numerical data are burned to DVDs and shelved with other data. Any remaining water samples are put in cold storage.

2.2 CENS Data, Users, and Uses

Data from CENS' dynamic field deployments can be grouped into four types. Sensors are used to capture numerical data on 1) the scientific application, 2) the performance of the sensors themselves, and 3) proprioceptive data to use in navigation for robotic sensor technology. The fourth category is hand-collected data for the scientific application, such as the water samples described above in the deployment scenario. Each of the four data types has multiple variables; from temperature and barometric pressure to roll, pitch, and yaw to packets sent and received. Some data serve only one purpose, but most serve multiple purposes for instance the scientific data collected can be used to identify sensor faults [9].

When we asked our subjects about capturing, using, sharing, and preserving data from deployments, and about capabilities they desired in archives to support their data, the primary interest was in the scientific data. Computer science and engineering researchers were as concerned about the quality and accessibility of scientific data as were the domain scientists. Conversely, the computer science and engineering researchers took little interest in maintaining access to sensor performance data or proprioceptive data that are essential to their own research. These forms of data appear to serve transient purposes for these researchers, with minimal archival value. However they may be essential for re-use of the application science data by others.

3. METHOD

Our research questions address the initial stages of the data life cycle in which data are captured and subsequent stages in which the data are cleaned, analyzed, published, curated, and made accessible. The interview questions were divided into four categories: data characteristics, data sharing, data policy, and data architecture. This paper reports our further exploration of the scientific data life cycle based on responses to questions about data characteristics and architecture. Findings on other questions are reported elsewhere [10-16].

The findings reported here are drawn from an interview study of five environmental science projects and subsequent field

observations. For each project, we interviewed a complementary set of science and technology participants, including faculty, post-doctoral fellows, graduate students, and research staff. CENS is comprised of about 70 faculty and other researchers, about 120 student researchers, and some full-time research staff who are affiliated with the five participating universities. Our pilot ethnographic study consisted of in-depth interviews with two participants, each two to three hours over two to three sessions. The intensive interview study consisted of 22 participants working on the five ecology projects. Interviews were 45 minutes to two hours in length, averaging roughly 60 minutes.

The interviews were audio-taped, transcribed, and complemented by the interviewers' memos on topics and themes. Transcription totaled 312 pages. Analysis proceeded to identify emergent themes. We developed a full coding process using NVivo 2, which was used to test and refine themes in coding of subsequent interviews. This study used the methods of grounded theory [17] to identify themes and to test them in the full corpus of interview transcripts and notes.

In addition to the interviews, members of our research team have attended numerous field deployments with various CENS research groups. Participant observation was used to gain entree to these work-intensive daylong to weeklong data collection events. While assisting the researchers in their deployment of equipment and data capture we were able to discover the processes leading up to the deployment and the plans for use of the data captured.

User scenarios for how data were captured, processed, and published were extracted from the interview and observational data. These scenarios were used to construct a data flow model, including the data sources, level of derivation, and any computer programs or scripts that were used to transform the data. From the combined flows we were able to extract common procedures and generalize them across our participants. We then verified this life cycle model during our interactions with researchers after the interviews, or observations of their data collections efforts.

4. RESULTS

Described here are both the general lifecycle phases and the three data lifecycles present within CENS research. Initially our research lead us to model a unified data life cycle that applied to all of CENS research. This lifecycle model is described in more detail along with the charge to evaluate the model for generalizability across the center and to shine alight on some of the more entangled phases (DCC). In teasing out the variation present in two of the phases of the lifecycle the generalization broke. There was a clear distinction between what happened to the data created as a part of technology research and those data created when science and technology researchers worked together, and these were both different from the data created by scientific research. These three types of lifecycle are characterized by what the research is meant to accomplish, be that scientific research, technological research, or development of technology for science. This latter is when both science and technology researchers collect data in the field, with the technologists responsible for the equipment or systems being tested and the scientists responsible for collecting other data to make the sensor data worthwhile.

4.1 General Life Cycle Phases

We have identified nine stages that appear to be common to the CENS deployments studied, the researchers, and to the resulting

data. The order of the steps is not absolute, as some stages are iterative while others may occur in parallel. For instance, Phases 4-6 appear to happen concurrently, with new outliers emerging only when the data has been integrated

- 1) Experiment Design. The beginning of the data life cycle is the design of new experiments. CENS researchers design new experiments by reusing data from prior research.
- 2) Calibration and Setup. Before sensors are deployed, they are calibrated to known solutions or values to identify the offset between the actual measurement and the expected measurement. They are calibrated again in-field, a process referred to as "ground-truthing".
- 3) Capture or Generation. Once sensors have been deployed successfully in the field, researchers begin to collect observations of physical phenomena. Some sensor measurements are direct (e.g., temperature, wind speed) and others are indirect (e.g., measure of fluorescence as an indicator of chlorophyll activity).
- 4) Cleaning. After data have been captured, calibration and ground-truthing information need to be applied to the data to normalize any calibration offsets from the sensing equipment.
- 5) Integration. Few of the observations and samples collected in the field can be interpreted without derivation into more meaningful data points. Data typically must be averaged into composite points before they can be used in analysis.
- 6) Derivation. Researchers are looking for trends over time and across spatial locations. Datasets each given deployment are integrated by multiple researchers, for multiple reasons, and in multiple combinations.
- 7) Analysis. Researchers use statistical, modeling, and visualization tools that vary by research specialty and individual preference. They test and generate hypotheses and draw conclusions about data obtained from the deployments.
- 8) Publication. Data collected during embedded network sensor deployments culminate in scholarly publications such as journal articles, conference papers, posters, and technical reports.

Preservation. Few, if any, of the CENS researchers interviewed had data preservation strategies commensurate with those of the archival community. It is more accurate to say that they back up their data.

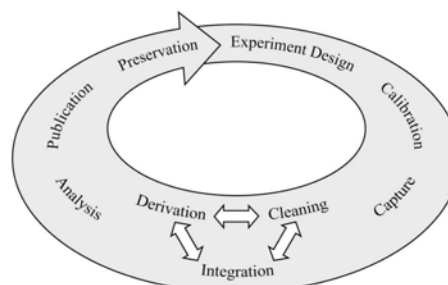


Figure: Life cycle of CENS data.

4.2 Three Lifecycles Model

Table: Three lifecycles with sample tasks for each phase of each cycle

Phase/Cycle	Scientific Research	Sci-Tech Development	Tech Research
<i>Experiment Design</i>	Generate hypothesis; develop methods; choose equipment; plan sampling schedule	Negotiate researchable questions; choose equipment and personnel; schedule tasks	Generate hypothesis; develop methods
<i>Calibration/Setup</i>	Calibrate equipment; collect ground truth samples	Calibrate sensing systems; ground-truthing	Prepare model, data, or algorithm to be used
<i>Data Capture/Generation</i>	Hand sampling; observation; processing samples	Sensor collection; hand sampling; observing environment; tweaking systems; observing users; checking in across groups	Sensor collection; generating from models; creating new data by running algorithms over data; creating models from data
<i>Cleaning, Integration, & Derivation</i>	Analysis of samples; recording presence and frequency/volume of organisms or chemicals; comparing to environmental models; remove outliers	<i>Part I: Tech</i>	Debug; investigate error reports; retesting; pass code around to get additional opinions
		Remove sensor artifacts; synch time stamps; recalibrate; aggregate data by variable; derive data for compound measures	
		<i>Part II: Science</i>	
		Sample analysis; recording presence and frequency/volume of organisms or chemicals; ground-truthing based on hand samples; comparing to environmental models; removing outliers	
<i>Analysis</i>	Linear regression of variables captured; hypothesis testing	Visualization; hypothesis testing	Comparisons; regressions; evaluation
<i>Publication</i>	Publish conclusions in science journals; post or deposit data	Publish conclusions in science journals and technical proceedings; post data	Publish conclusions in technical proceedings; post data
<i>Preservation</i>	Refrigerate samples; numerical data kept in databases; printed for hard copies; filed	Refrigerate samples; numerical data kept in databases; move files to a lab server or local machine	Move files to a lab server or local machine

In these models we have combined the formerly separate Cleaning, Integration, and Derivation phases because they are so interrelated. In order to properly remove outliers, the data must be derived or integrated, thus these three phases are really separate tasks that are performed iteratively until the data is clean enough for analysis to begin.

5. DISCUSSION

Technology research has the luxury of not always needing to go out into the field. Much of their research involves testing models or algorithms, generating data from these, and evaluating the performance, none of which requires fieldwork. At the same time the data generated in the lab may be wildly different from those captured in the field, and the occasional reality check is necessary to maintain a clear heading. When the technologists perform fieldwork, they do not need the scientists present to capture data that will progress their research, because their test subject in these cases are the equipment and systems themselves. Similarly the scientists do not need to bring the technologists along for their own data collection efforts, unless they want to use the sensing equipment.

Many Technology research cycles happen between in-field deployments, just as many in-field technology deployments happen between a Science-Technology Development deployment. For scientists on the other hand the frequency of data collection cycle is much lower, with most cycles lasting a year or more.

These science research cycles carry a much higher investment for the data collected. The interplay of the cycles are negotiated based on time available and need. Coordinating larger deployments to collect data that is meaningful to both the scientists and the technologists is a significant investment of time and resources. Difficult as they are these are the rare opportunities to see the "users" using the instruments and the systems, allowing the technologists to begin the next iteration of the design.

Comparing these three data lifecycles reveals the affects of collaboration on routine research by either the science or technology researchers. The experiment Design phase is much different during the collaborative cycle, attempting to offer significant research opportunities to both parties. The data capture phases are very different for each cycle; during the science research cycle the emphasis is on hand collected samples, the collaborative cycle is concerned more with the capture of sensor data than the hand sample data even though this data is significant to the science researchers and acts as a ground truth mechanism for the sensor data, and the technology research cycles range from collecting sensor data to generating data from a model. The cleaning, integration, and derivation phase is similarly different across all three cycles, but in this case the significant part is the way the custody of the data is no longer shared. The technology researchers must first make sense of the captured data in reference to the technology and then the science researchers will make sense of them in reference to the science. The final phase to display

divergence is the publication phase, where collaboration forces both the scientists and the technologists to publish in one another's domains. These publications do not carry the same weight as those within their own domain, thus dis-incentivising collaboration.

These affects caused by collaboration appear to be trouble spots where the future value of the data is compromised, due to variations in methods and processes across collaborating disciplines. For instance, the use of linear regression for data analysis does not scale to the volume of sensor-captured data, thus scientists must rely on technologists for new methods of interpreting their data. Linear regression is the established method for ecological data analysis, and has been so for multiple decades. The science researchers in this case are really dependent on the tools being built for them to be as trustworthy as linear regression and stand up to peer-review. Another example of this is the variation in what is considered data by each group, for instance the technology researchers will consider their algorithm or the script developed to be data, whereas the scientists will consider the hand sample or the sensor readings to be data. This may opens gaps where one party is not as invested in the overall quality of a specific data type because it is not central to their own discipline.

Actions taken at each stage of the life cycle influence how the resulting data can be interpreted, hence it is important that these stages be documented and associated with the resulting dataset. There is a cumulative effect of decisions made at each stage of the life cycle. For example, decisions made in the experimental design stage determine what data exist for analysis, or calibration decisions are essential to interpreting the data. There is a delicate balance of decision making between scientific and technology research partners.

6. CONCLUSIONS & FUTURE RESEARCH

The success of eScience depends upon successful collaboration between application scientists and their partners in computer science and engineering. Data resulting from such collaborations is expected to be extremely valuable for reuse by others. However, the value of data for reuse depends upon the quality of those data, which in turn depends on the ability to interpret the origins, provenance, and context of the data. Surprisingly little is known about how data arises from eScience collaborations. Our case study of ecological research in the Center for Embedded Networked Sensing sheds light on the various life cycles of eScience data.

Our future research will continue to explore and refine the data life cycle identified here, and to build systems to support it. In order to determine how generalizable this three-cycle model is across the various research groups, we need to test this model in the field. Additionally we would like to understand the role of publication in collaboration, specifically if the willingness to co-author is skewed towards either the science or technology researchers by exploring whether scientists are more apt to publish in technology papers, or the reverse. Given our access to the co-authorship data from this population answering this question is a logical next step. We are also curious about the role of decisions made in the field on the calibration and capture interplay, as well as the experiment and design interplay. At present, much of the sensing technology is experimental, but commercial off-the-shelf sensors also are in use. Research questions about data provenance will evolve as the technology stabilizes and the scientific research questions broaden.

7. IMPLICATIONS FOR iSCHOOL

Our interest in understanding information from every angle should lead us to studying the processes and methods that lead to the creation of data. Scientific data represents the entire evidentiary basis for scientific information and knowledge. Thus without studying the life cycle of data we are ignoring the premise of scientific information. Research such as this cuts to the heart of the disciplinary tributaries that comprise iSchool, namely Informatics, Archives, and Digital Librarianship, while also falling under Science, Technology, and Society. This is the practical application of learnings from each of these fields and serves as an example of what will advance iSchool education and research agenda.

8. ACKNOWLEDGMENTS

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. CENSEI, under which much of this research was conducted, is funded by National Science Foundation grant #ESI-0352572, William A. Sandoval, Principal Investigator and Christine L. Borgman, co-Principal Investigator. Alberto Pepe's participation in this research is supported by a gift from the Microsoft Technical Computing Initiative.

9. REFERENCES

- [1] Esanu, J.M., et al., Selection, Appraisal, and Retention of Digital Scientific Data: Highlights of an ERPANET/CODATA Workshop. *Data Science Journal*, 2004. 3.
- [2] Beagrie, N., Digital curation for science, digital libraries, and individuals. *International Journal of Digital Curation*, 2006. 1(1).
- [3] Beagrie, N. and D. Greenstein, A Strategic policy framework for creating and preserving digital collections. 1998, London: Arts and Humanities Data Service: London.
- [4] Digital Curation and Preservation: Defining the research agenda for the next decade, in Report of the Warwick Workshop - 7 & 8 November 2005. 2005, Digital Curation Centre: Warwick, UK.
- [5] Lord, P. and A. Macdonald, E-Science Curation Report--Data Curation for E-science in the UK: An Audit to Establish Requirements for Future Curation and Provision. 2003, JISC Committee for the Support of Research. p. 85 pages.
- [6] Day, M., Metadata for digital preservation: an update. *Ariadne*, 1999(22).
- [7] Geosciences Network. Last visited 16 August 2006 <http://www.geongrid.org/>.
- [8] National Ecological Observatory Network. Last visited 3 October 2006 <http://neoninc.org/>.
- [9] Ni, K., et al., Sensor Network Data Fault Types. *ACM Transactions on Sensor Networks*, in review.
- [10] Wallis, J.C., et al. Know Thy Sensor: Trust, Data Quality, and Data Integrity in Scientific Digital Libraries. in 11th European Conference on Digital Libraries. 2007. Budapest, Hungary: Berlin: Springer.
- [11] Borgman, C.L., J.C. Wallis, and N. Enyedy. Building digital libraries for scientific data: An exploratory study of data

FULL PAPER SUBMISSION 26:

Taylor's Value-Added Model: Still Relevant After All These Years

AUTHOR(S):

Eisenberg, Michael B
Dirks, Lee

ORGANIZATION(S):

University of Washington, United States of America
Microsoft Corporation

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Taylor's Value-Added Model: Still Relevant After All These Years

Michael Eisenberg, University of Washington
Lee Dirks, Microsoft Corporation

Abstract

Taylor's Value-Added model (1986) was a broad and ambitious effort to provide a unified framework for focusing on user needs and preferences in evaluating and designing information systems. Although developed in the early 1980s—before the wide-spread adoption of the PC, and well-before the Internet and Web-based technologies that have so changed our lives—the model holds up remarkably well in terms of explaining why and how various systems and systems attributes are useful and desirable or not.

The Value-Added Model seeks to explain what users want, why they want them, and how systems are able to meet (or not meet) those needs? “What do users want from information systems that would enable them to perform better, however “better performance” is defined in their context?” (Taylor p. 55) This paper updates Taylor's work in light of dramatic developments in information technology over the past 20 years and demonstrates how the model remains highly-applicable and valuable in both research and practical contexts across the interests of schools.

Bob Taylor is well-known for his contributions to the information field. His 1968 paper, “Question Negotiation and the Reference Process,” was one of the first works to emphasize a user and information perspective. It remains one of the most cited works in the history of library and information science. Taylor was also a visionary and pioneer in the movement that led to the formation of information schools. In the mid-1970s, he assumed the deanship at Syracuse, changed the name to the School of Information Studies and launched their doctoral program and later the Master's in Information Resources Management. Taylor finished his career with his work on the Value-Added Model.

As noted, the purpose of the Value-Added Model is to provide a framework for considering information and systems from a user perspective. Underlying the model are the three foundation elements of the information field—people, information, and technology:

1. People: The main focus is on the user. Systems exist to meet the information needs of users. Additionally, people can be viewed as part of the system.
2. Information: There is a hierarchy of information - the “information spectrum.” As value is added, we move up the spectrum from data to information to knowledge to action.
3. Systems: The purpose of an information system is to add value to better meet user needs. Various systems' processes add value in order to meet user needs.

The core framework underpinning Taylor's Value-Added Model is Figure 4.2 from his book (Taylor, 1986 p. 50) reproduced here:

USER CRITERIA OF CHOICE	INTERFACE (Values Added)	SYSTEM (Value-added Processes: Examples)
Ease of Use	Browsing	Alphabetizing
	Formatting	Highlighting important terms
	Interfacing I (Mediation)	
	Interfacing II (Orientation)	
	Ordering	
	Physical Accessibility	
Noise Reduction	Access I (Item identification)	Indexing
	Access II (Subject identification)	Vocabulary Control
	Access III (Subject summary)	Filtering
Quality	Accuracy	Quality control
	Comprehensiveness	Editing
	Currency	Updating
	Reliability	Analyzing and comparing data
Adaptability	Closeness to problem	Provision of data manipulation capabilities
	Flexibility	Ranking output for relevance
	Simplicity	
	Stimulatory	
Time-Saving	Response Speed	Reduction of processing time
Cost-Saving	Cost-saving	Lower connect-time price

(from Taylor, p. 50; Figure 4.2. User Criteria and Values Added)

“USER CRITERIA OF CHOICE” (first column) are the broad criteria categories that are important to users in choosing a system or in evaluating how well a system performs. These criteria are not absolute—the relative priority of one or another will depend on the user, situation, needs, etc. From experience in presenting and using the value-added model, we find the heading “user criteria of choice” to be sometimes confusing. Therefore, in our proposed revised model, we call this column, “User Value Criteria.”

The second column, “INTERFACE (Values-Added)” are the more specific values that are added in order to best meet the User Value Criteria. For example, accuracy, comprehensiveness, currency, reliability, and validity all can contribute to meeting the criteria of “quality.” Again, from experience, the original heading can be confusing. Also, we find it's not just about the interface. Therefore, in our revised model, we use “Values-Added” for this column.

The last column refers to the SYSTEM – the processes of the system that add the values in columns 2 and 1. The processes of quality control, editing, updating, and analyzing contribute to the values-added of accuracy, comprehensiveness, currency, reliability, and validity. Again, for clarity purposes, in our revised model, we refer to this column simply as “System Processes.”

Therefore, our first set of revisions to Taylor includes an edited set of overarching terminology:

USER VALUE CRITERIA -- VALUES ADDED – SYSTEM PROCESSES

Other more specific revisions are explained in more detail in the full paper. They include an additional criteria in the first column and further articulation of values-added (column 2) and system processes (column 3), namely tangible examples from today’s information ecosystem to clearly demonstrate the application of the framework to the real-world activities and scenarios. Indeed, the bulk of the paper demonstrates the widespread applicability of the model in current contexts to better understanding users, information, and systems.

The modified Taylor model helps explain the motivation of users, why certain systems and systems features perform so well in meeting user’s needs or not (e.g., electronic spreadsheets, email, Google, Amazon, GUI, the Web, social networks). Indeed, we posit that Taylor’s model can (and should) help to guide systems design, user studies, marketing, and entrepreneurship in information management. This last area may be its most compelling use. Entrepreneurs seeking to determine new products and services can utilize this updated Taylor model as a check-list for improving, enhancing or developing new and more compelling information products and services. In our view, the modified Taylor value-added model offers much as a model to understand and explain entrepreneurship.

The paper closes with an outline for further development, application, and research of the Taylor model. The ischool community continues to seek ways of explaining to wider audiences what it is that we do and why it is important. We believe that in re-acquainting the field with an evolved/updated view of Taylor’s seminal work, a functional model will greatly facilitate this important effort.

Authors:

Michael B. Eisenberg
Dean Emeritus and Professor
The Information School of the University of
Washington
Box 352840, Seattle, WA 98195-2840
Phone: (206) 616-1152
mbe@u.washington.edu
<http://www.ischool.washington.edu/mbe>

Lee Dirks
Director, Scholarly Communication
Technical Computing / External Research
Microsoft Corporation
Redmond, WA 98052
(425) 703-6866 - office
ldirks@microsoft.com
<http://www.microsoft.com/science>

Taylor's Value-Added Model: Still Relevant After All These Years

Mike Eisenberg, University of Washington
Lee Dirks, Microsoft Corporation

iConference
February 27-March 1, 2008
UCLA, Los Angeles, CA.

Introduction

This paper is an effort to reacquaint the information field with the work of one of its pioneers: Robert S. Taylor and his Value-Added Model. Taylor's Value-Added model (1986) was a broad and ambitious effort to provide a unified framework for focusing on user needs and preferences in evaluating and designing information systems. Although developed in the early 1980s—before the wide-spread adoption of the microcomputer, and well-before the Internet and web-based technologies that have so changed our lives—the model holds up remarkably well in terms of explaining why various systems and systems attributes are useful and desirable or not.^{*}

The Value-Added Model seeks to explain what users want, why they want them, and how systems are able to meet (or not meet) those needs? “What do users want from information systems that would enable them to perform better, however “better performance” is defined in their context?” (Taylor p. 55) This paper updates Taylor's work in light of dramatic developments over the past 20 years and demonstrates how the model remains highly applicable and valuable in both research and practical contexts across the interests of ischools.

Robert “Bob” Taylor is well-known for his contributions to library and information science. His 1968 paper, “Question Negotiation and the Reference Process,” (Taylor 1968) was one of the first works to emphasize a user and information perspective. It remains one of the most cited works in the history of library and information science.[†] Taylor was also a visionary and pioneer in the movement that led to the formation of information schools. In the mid-1970s, he assumed the deanship at Syracuse, changed the name to the School of Information Studies and launched their doctoral program and later the Master's in Information Resources Management. Taylor finished his career with his work on the Value-Added Model.

The goals of this paper are:

- (1) To reintroduce the field to Taylor's model.
- (2) To suggest revisions to the model based on our experience and our interactions with information professionals and graduate students.
- (3) To demonstrate the widespread applicability of the modified model in current contexts to better understanding users, information, systems, as well as the scope of the information field.
- (4) To offer recommendations for further work to develop and use the modified model.

^{*} We state this from personal experience in using Taylor's model in formal presentations and graduate courses.

[†] For example, a quick “Cited Reference Search in the ISI Web of Knowledge notes 255 citations for the 1968 College & Research Libraries paper.

Eisenberg/Dirks 2008 p. 2

The Taylor model (both the original and our proposed modified model) helps explain the motivation of users, why certain systems and systems features perform so well in meeting user's needs or not (e.g., electronic spreadsheets, email, Google, Amazon, GUI, the Web, social networks). Indeed, we posit that Taylor's model can (and should) help to guide systems design, user studies, marketing, and entrepreneurship in information management. This last area may be its most compelling use. Entrepreneurs seeking to determine new products and services can utilize this updated Taylor model as a check-list for improving, enhancing or developing new and more compelling information products and services. In this paper, we offer the modified Taylor value-added model as a means to better understand and explain successful entrepreneurship and innovation.

The paper closes with an outline for further development, application, and research of the Taylor model. The ischool community continues to seek ways of explaining to wider audiences what it is that we do and why it is important. We believe that in re-acquainting the field with an evolved/updated view of Taylor's seminal work, a functional model will greatly facilitate this important effort.

Taylor's Value-Added Model

As noted above, the purpose of the Value-Added Model was to provide a framework for considering information and systems from a user perspective. Underlying the model are the three foundation elements of the information field—people, information, and technology:

1. **People:** The main focus is on the user. Systems exist to meet the information needs of users. Additionally, people can be viewed as part of the system.
2. **Information:** There is a hierarchy of information - the "information spectrum." As value is added, we move up the spectrum from data to information to knowledge to action.
3. **Systems:** The purpose of an information system is to add value to better meet user needs. Various systems' processes add value in order to meet user needs.

Taylor emphasized that information systems are all about meeting the needs of users. Systems and the underlying system processes, algorithms, and features exist to add value in order to meet those needs. The Value-Added Model provides an organized framework for considering system processes that add value in order to meet user needs. Taylor's original Value Added framework is presented in Figure 1 (**Figure 4.2** from his book (Taylor, 1986 p. 50).

Eisenberg/Dirks 2008 p. 3

Figure 1: Taylor's Value-Added Model. From Taylor 1986, Table 4.2. p. 50.

<u>USER CRITERIA OF CHOICE</u>	<u>INTERFACE (Values Added)</u>	<u>SYSTEM (Value-added Processes)</u>
Ease of Use	Browsing Formatting Interfacing I (Mediation) Interfacing II (Orientation) Ordering Physical Accessibility	Alphabetizing Highlighting important terms
Noise Reduction	Access I (Item identification) Access II (Subject description) Access III (Subject summary) Linkage Precision Selectivity	Indexing Vocabulary control Filtering
Quality	Accuracy Comprehensiveness Currency Reliability Validity	Quality control Editing Updating Analyzing and comparing data
Adaptability	Closeness to problem Flexibility Simplicity Stimulatory	Provision of data manipulation capabilities Ranking output for relevance
Time-Saving	Response Speed	Reduction of processing time
Cost-Saving	Cost-saving	Lower connect-time price

The first column on the left, "USER CRITERIA OF CHOICE" includes the broad categories of criteria that are important to users in choosing a system or in evaluating how well a system performs. These criteria are not absolute or fixed. Consider the different situations of a senior NASA scientist and a 4th grade student. If both were using information systems to seek information about climate change in the Arctic, the scientist might rate quality (with the associated values of currency, accuracy, and reliability) as the top priority. For the 4th grader, ease of use (with the value accessibility) or cost-saving might be as if not more important. The relative priority of one or another criteria will depend on the person, situation, needs, setting, and other user-centered aspects.

The second column, labeled "INTERFACE (Values Added)" includes the more specific values that are added in order to best meet the USER CRITERIA OF CHOICE. For example, accuracy, comprehensiveness, currency, reliability, and validity all can contribute to meeting the user

criterion "Quality." The user criteria "Noise Reduction" relates to values of access, linkages, precision, and selectivity.

Taylor's last column is labeled, SYSTEM (Value-added processes). These are the processes, features, and elements of the system that add to the related values identified in column 2 (which in turn meet the user criteria of column 1). For example, the processes of quality control, editing, updating, and analyzing may contribute to the values added of accuracy, comprehensiveness, currency, reliability, and validity which then combine to address the user criterion of Quality.

As pointed out in the introduction, this model was developed well before many of the technological changes that have fundamentally altered human society, e.g., the personal computer, cell phones, the Internet, the World Wide Web. However, the model is robust and highly useful in explaining why these and other technological innovations are adopted and valued by individuals and organizations.

Taylor explains the intricacies of the model and defines various terms in Chapter 4 of his 1986 book. He also provides a table of definitions of his identified Values-Added. Rather than replicate Taylor's elaboration here, this paper first presents suggested modifications that clarify and expand the original Value-Added Model. This is followed by an abbreviated discussion of user criteria, values added, and system processes within the context of a suggested modified Value-Added Model.

Eisenberg-Dirks Modifications to Taylor's Value-Added Model

The core of Taylor's model is represented in Figure 4.2 from his 1986 book, reproduced above as Figure 1. Our suggested modifications relate to this figure and are presented below in Figure 2. While we have shared these modifications previously with various audiences through presentations, this is the first recorded paper outlining our thoughts. Therefore, we see these as formative or proposed modifications, and we expect that feedback from readers as well as from our field-based investigations will help us to fashion a more complete and conclusive Modified Value Added Model. In addition, we recognize the desirability of identifying, analyzing, and comparing frameworks and models of fundamental concepts of information, systems, services, and behaviors (e.g., relevance, credibility, use). We expect that this too will lead to adjustments in specific elements included in the modified model. For example, we anticipate that advances in the application of semantic technology could have major implications in the User Criteria of "Ease of Use" and "Noise Reduction."

Eisenberg/Dirks 2008 p. 5

Eisenberg/Dirks Modified Taylor's Value-Added Model, Feb 2008

Based on Table 4.2. Taylor 1986, p. 50.

<u>USER CRITERIA</u>	<u>VALUES ADDED</u>	<u>SYSTEM PROCESSES</u>
Ease of Use	Browsing Simplicity Mediation Orientation Ordering Accessibility	Alphabetizing Highlighting Formatting Simplifying
Noise Reduction	Item identification Classification Summarization Order Referral Precision Selectivity Novelty	Indexing Controlled vocabulary Filtering Selection Hyperlinking Semantic connecting Search
Quality	Accuracy Comprehensiveness Currency Reliability Validity Authority	Quality control Editing Updating Analyzing Selecting
Adaptability	Contextuality Flexibility Simplicity Privacy	Data manipulation capabilities Sorting Customizing User profiling Informed consent Choice
Performance	Time saving Cost saving Security Safety	Bandwidth Parallel processing Server size Processor speed Resource allocation/sharing Multi-tasking Common protocols, business practices Encryption Password protection
Pleasing	Aesthetics Entertaining Rewarding Engaging Stimulating	Design Interactive Gaming Reinforcing

Eisenberg/Dirks 2008 p. 6

Our first recommended changes to Taylor's original model relate to the overarching terminology used. Taylor presents a three part view as the basis for the model: user, interface, and system. The "User" part focus on the problems which establish the criteria of choice. The "Interface" is a negotiating space between the user and the system and, according to Taylor, displays the values added by the system to assist the user in making choices. The "System" includes the specific processes that add specific values. (Taylor 1986, p. 49)

These labels can be difficult in explaining and applying the model, partially because of using the word "values" and term "value-added" in slightly different ways in the same table. We therefore offer a simplified labeling of the three part view:

- User Criteria
- Values-Added
- System Process

Instead of "user criteria of choice" it's more direct to simply state, "User Criteria." For column two, "Values-Added" encompasses much more than just the interface. The term "Values-Added" is still consistent with Taylor's description of these being the "values added by the system which aid customers in matching their needs." (Taylor 1986, p. 51). Lastly, in his text, Taylor refers to value-added processes and system processes interchangeably. "System processes" is preferred here as it avoids confusion with the "Values-Added" from column two. The use these terms helps to emphasize the relationships across the columns of the model. That is, various system processes.

Our second group of modifications involves the more specific elements of the model in each of the three columns. Again, while the original Value-Added Model is still useful for describing and analyzing current user and system interactions, developments over the past twenty years point to additional criteria and related values added. Taylor himself notes that his criteria are broken into six categories for convenience sake and are not graven in stone. (Taylor 1986 p. 51) We agree with his statement that they are a useful way of organizing the values that are added, but we offer a reformulation of the last two criteria (Time-saving and Cost-saving) under the broader heading of Performance. We also propose an additional criterion, Pleasing, which relates to the values of aesthetics, entertaining, reward, and engaging.

Taylor sought to derive his criteria and values added from the literature, but found it equally important to learn from professional experience. He found few studies concerned directly with values and user benefits, so he drew on various formal and informal sources. At this point in time, our suggested modifications in criteria, values added, and system processes derive from personal reflection and field-based experience. We recognize the importance of comparing and testing the modified model in relation to the literature as well as through systematic investigations.

Turning to column two, now labeled simply "Values Added," Taylor's original list is both consolidated and expanded. Here are some examples:

- For user criterion, "Ease of Use," Taylor offers mediation and orientation as two different ways of "interfacing." The modified model focuses on each more separately with the value of "Mediation" referring to assistance/help and "Orientation" for user familiarity or navigation with the system. The modified model also recognizes that the value of "Accessibility" is not just limited to physical, but virtual as well.

- For “Noise Reduction,” the revised model seeks to break out three values added nested within the term “access.” The modified model emphasizes item identification, classification, and summarization as separate values added while also offering “Referral” and “Novelty” as other values added to include.
- “Cost saving” and “time saving” were criteria only minimally developed in Taylor’s original conceptualization. Cost saving was also listed as both a criterion and a value. In the modified model, both of these are now considered as Values related to the broader criterion of Performance. We expect further development of this important criterion in terms of associated values and system processes.
- “Security,” “Privacy,” and “Safety” are three identified values added that are increasingly important. Security and Safety were determined to best relate to the newly combined user criterion of “Performance” while Privacy seemed increasingly associated with user-driven desires and therefore best associated to “Adaptability.”
- As noted above, “Pleasing” is a totally new user criterion, with associated values added of “Aesthetics,” “Entertaining,” “Rewarding,” and “Engaging.”

In certain instances, the values identified in “Values Added” can be applied to different “User Criteria.” For example, the value “Simplicity” relates to both Ease of Use and Adaptability. In addition, the terms “relevance,” and “usefulness,” are broad, multidimensional concepts of information and relate to many of the User Criteria. Relevance and usefulness may be viewed as infused across the *entire* Modified Value-Added Model. In the 1986 book, Taylor included a table of “Definitions of Values Added (see page 69) for all of the values included in the middle column. We hope to be able to do the same in a subsequent paper.

The third column, “System Processes,” provides examples of the capabilities, approaches, and features of systems that might be used to add the values (noted in column 2) to meet the criteria of users (column 1). Systems processes are numerous and continually being developed. Taylor’s original model only includes a small number of system processes. We attempted to identify more, however even these are only representative of abundant options.

For example, although likely inferred in Taylor’s original work, the concept of navigating connections via electronic links (labeled as the system process “Hyperlinking” in the Modified Model) is now obvious and a critical concept for nearly all information systems. Hyperlinking was conceived by computer scientist Ted Nelson in the 1960s, but it didn’t gain widespread use until the development of the World Wide Web by Tim Berners-Lee decades later. Still, it is not an exaggeration to say that this relatively recently-implemented system process—hyperlinking—is one of the most significant of all system processes because it is the foundation of the World Wide Web, our most pervasive and now essential information system.

The Modified Value-Added Model helps to explain why this is so. Hyperlinking is a System Process that relates to the values added of “Item Identification,” “Referral,” “Precision,” and “Selectivity.” These values, in turn, help to meet the User Criteria of Noise Reduction. This may seem strange at first because the World Wide Web is often criticized as noisy and contributing to information overload. However, that’s not due to hyperlinking. In fact, hyperlinking helps users

Eisenberg/Dirks 2008 p. 8

to sift through the noise by identifying items of interest, by referral to precise websites or sections of websites, or by allowing creation of selected links by website or content creators.

The usefulness of other information systems are similarly explained by the model by making the connection to related Values Added in order to meet identified User Criteria. Consider cell phones as an information system. For some people when choosing a phone, it's the "Pleasing" criteria that are most important. For these users it's important to emphasize the values added of aesthetics, entertaining or stimulating. These values are added by the systems processes of design and interactivity. For other users, it's "Ease of Use" that's crucial, and they are looking for a cell phone that is simple, accessible, and easy to navigate through functions (i.e., orientation). System processes of highlighting, formatting, and simplifying help to address these values and criteria.

For a final example, we return to the situation and needs of the senior NASA scientist and a 4th grade student noted earlier in this paper. Both were seeking information about climate change in the Arctic. For the scientist, it is likely to be all about "Quality" of information and the values that meet the quality criterion are accuracy, currency, comprehensiveness, authority, and reliability. System processes that address these values include quality control in research studies, editing and updating of papers and reports, analyzing data sets, and selecting valid and reliable information from authoritative sources. But, the 4th grade student has different needs that might relate to "Ease of Use," and also "Pleasing." System processes of formatting, highlighting, and simplifying may add values of simplicity, accessibility, and simplicity for ease of use. For pleasing, there are processes that provide interactivity and reinforcement.

Finally, although we anticipate the Modified Value-Added Model to evolve in time, we feel that there are varying degrees to which each column may or may not change. For example, the "User Criteria" column seems sufficient and complete and unlikely to change substantively. We made only minor modifications to the original concepts established by Taylor—combining Cost Effective and Time Saving into Performance and adding the new criterion, "Pleasing." The "Value-Added" column is somewhat definitive, but certainly not comprehensive. We expect this column will require some adjustment and expansion. The most frequent changes will take place in the "System Processes" column. As with Taylor, this column is intended to be representative and to change as new system processes and capabilities emerge.

Applying the Modified Value-Added Model

The sections above presented Taylor's original model and explained our proposed modifications. We also attempted to demonstrate how the model can be used to explain how different information systems can add value in order to meet user needs. In our view, the sustained relevance of Taylor's Value-Added work is impressive. His concepts and framework are as applicable and useful as they were twenty years ago. This section offers thoughts as to how the Modified Value-Added Model is applicable and valuable in practical and educational contexts.

Practical Applications

The role of information professionals has evolved dramatically over the past 20 plus years. In the ever-quickenning transition from the information field's past (e.g., punch and catalog cards) to the future (e.g., XML and metadata), the information manager is now thrust into the forefront of a multi-billion dollar information technology industry. Librarians and other traditional information professionals are now competing with major corporations and an entire industry of

Eisenberg/Dirks 2008 p. 9

information service and system providers facilitating access to information. This explosion of the information role has evolved and will continue to change, and in this process information professionals can continue status-quo, or can take charge and proactively adapt to address the growing information needs of end-users. In this new environment – this global, connected information marketplace with a new level of attention and scrutiny, it is critical for the information profession to take a new approach.

The development of information systems and the provision of information services in industry are fast-paced and ever-changing. It is an extremely aggressive marketplace, where competitive differentiation and viable substitutes are apparent daily. In this setting, the availability of a simple framework that can be used to brainstorm and assess potential ideas is an extremely valuable resource. Using the Modified Value-Added Model to evaluate and re-evaluate information systems allows the information professional a way to test and improve services and resources in terms of meeting users' needs. The consistent, focused application of this model will result in improved offerings, more successful systems and tools resulting in higher productivity. These types of benefits can have impact on all sectors, not just private industry. Shifting from the private sector, it is crucial that information professionals in the public sector utilize this framework as a tool as well. Whatever setting one is in, the new concept of "infopreneurship" is taking on greater significance.

For example, assume an aerospace engineer is working on a new project. For her, the most important User Criteria is Quality. An information professional seeking to assist this user could consider the values options noted in the middle column (e.g., Accuracy; Comprehensiveness; Currency; Reliability; etc.) and begin to think about how to add these values given the information set or the systems available. If the available systems are not able to meet the desired values, other systems or services can be brought into play. In this way, each of the first two columns of the Modified Value-Added Model serve as useful pivots that provide information professionals or managers a framework for considering possible products, services, or systems to use.

From an industry perspective, the Modified Value-Added Model can be applied in at least three ways:

1. **Developing New Systems & Tools:** If charged with developing a whole new information product or information service from scratch, the Modified Value-Added Model can provide a thorough, defined way of vetting potential ideas on top of whatever information source you are working with.
 - Scenario: Two business partners are considering the creation of a brand new online travel guide. They can consider the User Criteria of "Quality" and then analyze at the different Value Added options. It is important to keep in mind that each Value Added can be considered to have a broad range of choices:

Value Added: QUALITY	<i>High-End of the Quality range</i>	<i>Low-End of the Quality range</i>
Accuracy	Editors vetting or fact-checking data prior to publishing; established human quality-control process.	Automatically re-directing information feed from another source without review.
Comprehensiveness	Content is sourced from	Content only acquired from a

Eisenberg/Dirks 2008 p. 10

	multiple definitive sources, on a global basis.	single source (from the US).
Currency	Website content is updated real-time, as changes occur, updates are automatic.	Web-site content is updated manually, twice a week
Reliability	Links on the site are automatically checked nightly, and are deleted if not resolvable.	System doesn't permit automatic link checking, so this is a manual process.
Etc.		

Note: In the above example, automatic vs. manual is an important distinction between the two (fictitious) product offerings: if some cases, automatic is high-quality because it is instantaneous, whereas in other cases it can be low-quality since it has no human/editorial review. In this way, you can see the due-diligence that needs to go into the value-added assessment process.

2. Refining / Improving Existing Tools: Likewise, if the goal is to assess and improve and existing offering, this Model can help to develop scenarios to generate and test such enhancements. Following the scenario above, once the small firm's online travel website was doing well and the partners need to innovate to attract new customers or lure customer away from a competitor, they can look at possible ways to enhance your existing services.

- Scenario: Focusing on the "Adaptability," here are some possible ideas on how to brainstorm new ideas for the existing version of the travel website:

Value-Added: "ADAPTABILITY"	Potential Service Enhancement
Contextuality	Website knows if users are accessing the site from a non-US domain, and is able to high-light local/regional travel tips depending on your location.
Flexibility	Website knows if users are accessing via a PC or a mobile device, and automatically adjusts content to the appropriate screen-size.
Simplicity	If users repeat the same trip on a regular basis, website allows you to replicate the itinerary automatically to facilitate future travel.
Privacy	Users are able to login to a private section where they can save trip itineraries or bookmark favorite travel recommendations, etc.
Etc.	

Studying Competitive Offerings: Even if the immediate goal isn't to produce an offering, but instead to better understand and assess what directions other products and services might be taking, one could utilize this same approach with the intent of exhaustively listing out future developments by other players in the space. The key is to use the Modified Value-Added Model to systematically identify user needs and preferences, the values that will meet those preferences, and the system process that will add the desired values. The model facilitates prioritization by providing a framework linking user needs, values, and system processes as well as specific items within each.

For businesses and organizations, the Modified Value-Added Model can systematically assist in assessing and understanding:

- The landscape of **environment/market-space**.
- Varying **user/customer needs** and potential product/service requirements.
- Other **competitors and/or substitutes** that are present.
- Existing offering(s), to help **define new functionality** and why it should be **successful**.

Educational Applications

Information school graduates need to be prepared to enter a marketplace of ever-increasing complexity. Either as employees in a private sector business or a public sector organization (including libraries), all areas are now competing against new forces in our traditional realm—namely corporate entities such as Google or Microsoft—who are facilitating access to an ever-growing volume of information. It is critical that we prepare students to learn to vie for the time and attention of information-seekers, to evolve our approach. How that approach is manifest itself has changed dramatically in the past 10-15 years and will need to continue to evolve to keep pace with the developments in the marketplace as well as the needs of end users seeking information for their daily tasks.

Students benefit from a more systematic approach that includes an enhanced service-orientation and an overall greater entrepreneurial tact. Taylor's Modified Value-Added Model is a prime example of the sorts of tools and resources that need to be developed, tested, utilized and evangelized across our profession. Only in this way can we hope to ensure our students are well-positioned for success in the job market.

Over the past three years at the University of Washington, both authors have successfully used the Modified Taylor Model in both undergraduate and graduate-level classes including Foundations of Informatics, Human Aspects of Information Systems, and the Life Cycle of Information. We find that the model helps students to move from theoretical understandings of the "user perspective" to a something more tangible. Students are able to apply the model to a range of user, use, and systems situations including:

- Persons selecting a cell phone or mobile device.
- The merits and limitations of car dashboard layouts.
- Comparison of search engines.
- Students in distance learning courses.
- Sports information systems.

A typical class exercise, lab, or assignment related to the Value-Added Model requires students to be able to complete the following:

1. Identify an information situation involving users, needs, and information systems.
2. Describe the situation and use from a systems (input-process-output) perspective.
3. Describe the situation from a user perspective.
4. Analyze the situation in terms of:
 - User Criteria
 - Values Added
 - System Processes.

Eisenberg/Dirks 2008 p. 12

In every instance, using the Modified Value-Added model to analyze of users, situations, needs and systems result in rich discussion and brainstorming with students. Applying the model also seems to engage them to think in a more systematic and entrepreneurial manner.

Future Work

As noted at the outset, this paper seeks to reintroduce Taylor's Value-Added model because of its usefulness as a framework for focusing on user needs and preferences in evaluating and designing information systems. Through analysis, reflection, and explanation, we attempted to demonstrate how a Modified Value-Added Model remains highly applicable and valuable in practical and educational contexts.

In terms of research, there is considerable work to be done in terms of (1) further developing and refining the Modified Value-Added Model; (2) applying and testing the model across situations; and (3) using the model in research studies to better understand users, needs, and systems. In addition, we are particularly interested in (4) studying innovation and entrepreneurship from a value-added perspective. Lastly, we hope to (5) more systematically determining the usefulness and impact of the model in information school educational programs

Further development and refinement of the Modified Value-Added Model involves delving deeper in to the literature to make connections between the Value-Added Model and other conceptualizations of key concepts and understandings in information science (e.g., Saracevic (2007) and Schamber (1994) on relevance; Metzger (2007) on credibility). There is also literature to consult related to information behavior and information management. We also encourage others to share insights on this paper. We are considering setting up a Wikipedia entry on Taylor's Value-Added Model in order to facilitate interaction and discussion and to publicly and collectively further develop the model. We also encourage others to engage in follow-up studies and share results with us and through various forms of publication.

Our preference for applying and testing the model across situations involves field-based research in businesses, particularly information oriented start-up companies. We envision both qualitative (case study) approaches as well as qualitative data gathering related to users and situations in terms of User Criteria, Values Added, and System Processes. Here too, we hope that other researchers will consider using the model in user behavior and systems studies.

Development and research related to educational programs will involve working with faculty and students across information schools, programs, and courses. This paper is a first effort to share our thoughts and experiences with the broader information field and education community. Our hope is that this paper will serve to raise interest in using the model in courses and programs. If there is interest, we would be interested in forming an online community interested in sharing ideas, materials, and approaches. This might be followed by a systematic study of the usefulness and impact of the model in information education programs.

Eisenberg/Dirks 2008 p. 13

References

- Metzger, Miriam M. (2007) "Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research," *Journal of the American Society for Information Science and Technology*. Hoboken: Nov 2007. Vol. 58, Iss. 13; p. 2078
- Saracevic, Tefko. (2007) "Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance," *Journal of the American Society for Information Science and Technology*. Hoboken: Nov 2007. Vol. 58, Iss. 13; p. 1915
- Schamber, Linda. (1994) "Relevance and information behavior," In: *Annual review of information science and technology*, v29, 1994. Learned Information, 1994
- Taylor, Robert S. (1968) "Question Negotiation and the Reference Process," *College & Research Libraries*, 29 (3): 178-194 1968.
- Taylor, Robert S. (1986) *Value Added Processes in Information Systems*, Norwood, NJ: Ablex.
-

Authors:

Michael B. Eisenberg
 Dean Emeritus and Professor
 The Information School of the University of
 Washington
 Box 352840, Seattle, WA 98195-2840
 Phone: (206) 616-1152
mbe@u.washington.edu
<http://www.ischool.washington.edu/mbe>

Lee Dirks
 Director, Scholarly Communication
 Technical Computing / External Research
 Microsoft Corporation
 Redmond, WA 98052
 (425) 703-6866 - office
ldirks@microsoft.com
<http://www.microsoft.com/science>

FULL PAPER SUBMISSION 27:

FRBR as an interdisciplinary high-middle range theory

AUTHOR(S):

Renear, Allen H.

Dubin, David S.

ORGANIZATION(S):

University of Illinois at Urbana-Champaign, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Introduction

Science proceeds through the criticism, empirical and theoretical, of competing explanatory theories and hypotheses. Unfortunately within contemporary information science a sense of evolving scientific explanation is all too rare. Instead “brute empiricism” oscillates with vague generalities — the golden mean of Merton's “middle range theory” is apparently a hard target to hit when the problem space is interdisciplinary.

Perhaps a promising, if unintended, foray into that middle range is the Functional Requirements for Bibliographic Records (FRBR; IFLA 1998). FRBR is a “conceptual model of the bibliographic universe” (works, texts, editions, documents, authors, subjects, and the like) that was developed by the International Federation of Library Associations and Institutions. Its immediate objective is to guide the development of systems for creating and managing bibliographic records in order to better support the fundamental user tasks of discovery and use; it is not intended as a radical revision of existing practice or theory, but as an articulation current best practice the emerging theoretical consensus, with new terminology and refinements as needed.

Over the last few years the FRBR framework has been found natural and compelling and is increasingly reflected in cataloging practices and technology development in libraries and elsewhere: international bibliographic databases (such as Worldcat) and software systems (such as Endeavor) are being “FRBRized”, and the working group for the next revision of the bible of library cataloging, the Anglo-American Cataloging Rules (now Resource Description and Access) refers to FRBR as part of the “conceptual foundation” for that revision.

Although primarily developed for the library cataloguing community and focused on bibliographic record management FRBR is now increasingly recognized as a compelling empirically grounded framework for intellectual material in general, and is influential in other domains that are using ontologies for content management.

FRBR as a high-middle range theory

R. K. Merton has suggested that social science focus on “middle range” theories, rather than, on the one hand, mere hypotheses with little explanatory power, or, on the other hand, high-level all-encompassing theories that are currently beyond the reach of clear definition or empirical confirmation. We think this is a good recommendation for information science as well and we think that FRBR is a good candidate for what we would call a “high middle range” theory — surprisingly perhaps, considering its origins and objectives.

Since its original development FRBR has generated both practical applications and empirical investigations. However both of these would be common to lower level working hypotheses as well as middle range theories. And while such working hypotheses can be interdisciplinary in the sense of combining disciplines, they cannot demonstrate explanatory interdisciplinarity of the most interesting kind — because they are not in any deep way general explanations at all.

As interesting as ongoing empirical applications of FRBR are we think that even more revealing is the theoretical and formal analysis it has received. Such analysis is particularly important, perhaps even distinctive, for high middle range theories as it not only shapes the nature of hypotheses that are generated, but, more importantly, establishes relationships with other upper and middle range theories in the discipline, and, most importantly of all perhaps, establishes relationships with upper and middle range theories in other fields. In this way theoretical discussions of middle range theories typically reveal intrinsic connections across disciplines, not only exposing non-trivial interdisciplinarity but also providing the connections with established explanatory theories in other domains, which is widely recognized in the philosophy science as confirming the explanatory power, and warrant, of a candidate theory.

FRBR from a formal point of view

We review a number of theoretical discussions of FRBR as a formal model for bibliographic objects. In each case we note how the discussion surfaces rich connections with successful mature theories and analyses in other fields, revealing a web of interdisciplinary relationships that improve the scope and explanatory power of FRBR. The discussions covered include:

- -An argument for systematic reinterpretation of types as roles that connects FRBR closely to theories in pragmatics (speech act theory), formal ontology, social theory, and aesthetics. (Renear & Dubin 2006).
- -Treatment of possible confusions around concepts such as inheritance (Carsen et al. 2003; Renear & Choi 2006) and manifestation (Doerr, Hunter, & Lagoze 2003); this applies work from knowledge representation and semantic networks.
- -Suggestions for fundamental concepts that improve scope and explanatory power with additional causal relations; this applies recent work in philosophical logic. (Dousa 2007).
- -Accounts of entity assignment puzzles (Floyd & Renear 2007; Renear et al 2003).
- -Presentations of set theoretic (Svenonius 2001) and object-oriented (Doerr & LeBoeuf 2007) alternatives.

In addition to the insights provided into the explanatory nature of FRBR and the extent of the actual and potential interdisciplinary connections, it is also apparent that that these connections provide routes by which additional results from other fields can be integrated.

References

- Carsen, T., Garcia, H., Mabragana, C., Manzanos, N. (2003). Documentary description model based on the object paradigm. Proceedings VI Congreso del Capitulo Espanol de ISKO, Salamanca (Spain).
- Doerr, M., Hunter, J., and Lagoze, C. (2003). Towards a core ontology for information integration. *Journal of Digital Information*, 4(1).
- Doerr, M., LeBoeuf, P. (eds). (2007) FRBR: Object-oriented definition and mapping to FRBR-ER. Version 0.8. International Working Group on FRBR and CIDOC CRM Harmonisation.
- Dousa, T. (2007) untitled draft.
- Floyd, I., and Renear, A. H. (2007). What is the digital world is a FRBR item? Poster at The American Society for Information Science, October.
- Guarino, N. and Welty, C. A. (2000). A formal ontology of properties. In *Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management. Lecture Notes In Computer Science*. Springer-Verlag, 97-112.
- International Federation of Library Associations (IFLA). (1998). *Functional Requirements for Bibliographic Records: Final Report*. UBCIM Publications-New Series. Vol. 19, München: K.G.Saur, 1998.
- Levinson, J. (1980). What a musical work is. *The Journal of Philosophy*, 77, 5-28.
- Merton, R. K. (1968). *Social Theory and Social Structure*. The Free Press.
- Renear, A. H., Phillippe, C., Lawton, P., and Dubin, D. (2003). An XML document corresponds to which FRBR Group 1 entity? In *Proceedings of Extreme Markup Languages 2003*. Montreal, Canada, August.
- Renear, A. H., Choi, Y. (2006). Modeling our understanding, understanding our models -- the case of inheritance in FRBR. *Proceedings of the 69th ASIS&T Annual Meeting*, Austin, Texas.
- Renear, A. H. and Dubin, D. (2007). Three of the four FRBR Group 1 entity types are roles, not types. *Proceedings of the American Society for Information Science*.
- Searle, J. R. (1995). *The Construction of Social Reality*. New York: The Free Press.
- Svenonius, E. (2001). *The Intellectual Foundation of Information Organization*. MIT Press.
- Tillett, B. B. (2004). What is FRBR? A Conceptual Model for the Bibliographic Universe. Washington: Library of Congress, Cataloging Distribution Service. Online: <http://www.loc.gov/cds/downloads/FRBR.PDF>

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

Panel on Reading the Information Infrastructure

FRBR as an interdisciplinary high-middle-range theory
for information science — a theoretical perspective

Allen H. Renear
Graduate School of Library and Information Science
University of Illinois at Urbana Champaign
Champaign, IL 61820
+1 (217) 265-5216
renear@uiuc.edu

Dave Dubin
Graduate School of Library and Information Science
University of Illinois at Urbana Champaign
Champaign, IL 61820
+1 (217) 244-3275
ddubin@uiuc.edu

ABSTRACT

We suggest that IFLA's Functional Requirements for Bibliographic Records is an interesting, if unexpected, example of Merton's "theories of the middle range" and show how theoretical analysis and refinement of such theories can illuminate the deep interdisciplinarity of information science.

Topics

Cultural information systems, Information management, Information organization, Nature and scope of iSchools and iResearch

Keywords

FRBR, information science, R. K. Merton, theory, conceptual modeling, ontology.

58. INTRODUCTION

Science proceeds through the criticism, empirical and theoretical, of competing explanatory theories and hypotheses. This picture is may be an oversimplification, but it is a common enough scenario nonetheless.

Within contemporary information science however a sense of evolving scientific explanation, with theories generating hypotheses and undergoing both empirical and conceptual revision is still not as routine as one might like. Too often "brute empiricism"[110] seems to oscillate with vague generalities.

R. K. Merton has suggested that social science focus on "theories of the middle range", rather than, on the one hand, mere hypotheses with little explanatory power, or, on the other hand, high-level all-encompassing theories that can be neither clearly defined nor empirically confirmed.[102]

Although there are many promising middle-range theories in information science, there are not, we think, enough — the golden mean of Merton's middle range is apparently a hard target to hit when the problem space is interdisciplinary.

For our part in this panel we will discuss a surprising candidate for a middle-range theory in information science, IFLA's Functional Requirements for Bibliographic Records (FRBR).[99] Given its origins and objectives it may seem odd to describe FRBR as a theory in information science, but we have found doing so illuminating and have come to feel that regardless of original intention it is indeed a theory, and a good one.

Although guiding empirical research is the principal characteristic feature of middle-range theories, our discussion at this panel will take a theoretical rather than empirical perspective. This is an aspect of middle-range theories that is

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse

generally neglected, and it is one that we think nicely exhibits the deep interdisciplinarity of information science.

59. MIDDLE RANGE THEORIES

According to Merton theories of the middle range⁵

...lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization, and social change. [102] (p. 39)

It is characteristic of middle-range theories that they are not directly inferred from experience but rather themselves generate inferences about experience:

Each of these theories provides an image that gives rise to inferences. To take but one case: if the atmosphere is thought of as a sea of air, then, as Pascal inferred, there should be less air pressure on a mountain top than at its base. The initial idea thus suggests specific hypotheses which are tested by seeing whether the inferences from them are empirically confirmed. The idea itself is test for its fruitfulness by noting the range of theoretical problems and hypotheses that allow one to identify new characteristics of atmospheric pressure. [102] (p. 40)

These inferences "guide empirical inquiry":

Middle-range theory is principally used in sociology to guide empirical inquiry... it is intermediate to general theories of social systems which are too remote from particular classes of social behavior, organization, and change to account for what is observed, and to those detailed orderly description of particulars that are not generalized at all. [102] (p. 39)

Finally, middle-range theories are limited in scope:

Middle-range theories involve abstractions, of course, but they are close enough to observed data to be incorporated in propositions that permit empirical testing. Middle-range theories deal with delimited aspects of social phenomena ... One speaks of a theory of reference groups, of social mobility, or role-conflict and of the formation of social norms just as one speaks of a theory of prices, a germ theory of disease, or a kinetic theory of gases. [102] (p. 39)

Although middle-range theories are not, at least in the usual circumstances, derived from more general theories, Merton

⁵ Merton cites among his historical antecedents in commending middle-range theories Bacon (*axiomata media*) and Mill ("middle principles"), and indicates Durkheim's *Suicide* and Max Weber's *The Protestant Ethic and the Spirit of Capitalism*, examples of middle-range theories in social science.

does note that they may have logical relationships to those broader theories.

60. FRBR

The Functional Requirements for Bibliographic Records (FRBR) is a “conceptual model of the bibliographic universe” developed by the International Federation of Library Associations and Institutions to provide “a generalized view” of bibliographic entities and relationships.[99]

FRBR has as its immediate objective guiding the design of systems for creating and managing bibliographic records in order to better support the fundamental user tasks of discovery and use. It is not intended as a radical revision of existing practice or theory, but as an articulation of current best practice and an emerging consensus, with new terminology and refinements as needed.

FRBR has been very influential. Over the last few years the FRBR framework has been found natural and compelling and is increasingly reflected in cataloging practices and technology development in libraries and elsewhere: international bibliographic databases and software systems are being “FRBRized”, and the working group for the next revision of the bible of library cataloging, the Anglo-American Cataloging Rules (now Resource Description and Access) refers to FRBR as part of the “conceptual foundation” for that revision.[100]

FRBR divides bibliographic entities into three groups: Group 1 (the “products of intellectual and artistic endeavor”), Group 2 (their creators), and Group 3 (their subjects). We describe Group 1 in more detail, partly to give a sense of how FRBR is structured, but also because our example focuses on this group.

FRBR uses generic entity-relationship modeling techniques to express the formal features of the framework. The FRBR Group 1 entity types are works, expressions, manifestations, and items. A work is defined as “a distinct intellectual or artistic creation”, an expression is “the intellectual or artistic realization of a work in the form of alphanumeric, musical, or choreographic notation, sound, image, object, movement...”, a manifestation is “the physical embodiment of an expression

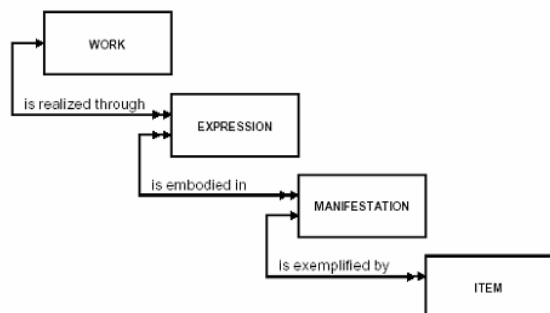


Figure 1. ER Diagram of FRBR Group 1 Entities (diagram from IFLA, 1998)

of a work”, and an item is “a single exemplar of a manifestation”.

Using printed books as an example (which we will do throughout) these concepts would correspond roughly to the common notions of work, text, edition, and physical copy, respectively.

Each entity type is associated with a characteristic attributes — for instance, works have form (novel, play, poem, etc.), expressions may be in a particular language, manifestations may have a typeface, and items may have a condition. A particular work may be realized in any number of expressions (such as different translations or textual variants); an expression may be embodied in any number of different manifestations (such as different editions with different page design or carrier); and a particular manifestation may have any number of individual physical instances. Works, expressions, and manifestations are abstract objects, and items are concrete physical objects.⁶

61. THEORETICAL ASPECTS OF MIDDLE-RANGE THEORIES

At the very heart of the notion of a middle range theory is the view that they guide empirical research by providing hypotheses for exploration, and by explaining empirically observed phenomena. A full account of FRBR as a middle range theory would therefore naturally focus on these hypotheses, the resulting research, and the effectiveness of the theory in explaining empirical observations. However this topic, as important as it is, and as timely as it is, will not be taken up in here. We focus on a different, and somewhat neglected, aspect of middle-range theories: the role of theoretical analysis and refinement in their conceptual evolution.

Merton has relatively little to say about the role of theoretical refinement in the function and evolution of middle-range theories. He does note that good middle range theories pose theoretical problems as well as guide research, and he remarks that while middle-range theories are not derived from upper range theories they may be consistent (and so also, presumably, inconsistent) with upper-range theories. He says little beyond that.

But it would seem that in the case of at least some theories, let’s call them upper middle-range theories, theoretical analysis and refinement is in fact a major force in their evolution, playing a distinctive role in how those theories provide scientific value, and, in particular, how they are integrated with other theories at higher levels, or at the same level, and, in particular, with theories in other fields. So as interesting and important as the ongoing empirical studies of FRBR are, we will for now ignore them entirely and focus here on theoretical and formal analysis. We hope to show, by an example, how such formal analysis establishes critical relationships with other upper and middle range theories both within information science, and across disciplinary boundaries. These sorts of integrating relationships among theories — vertical, lateral, and

⁶ For a short overview of FRBR see Tillett.[112]

interdisciplinary relationships — have often been identified as a source of enhanced explanatory power and warrant for scientific theories. And we think our example confirms this.

62. TYPES, ROLES, AND CONTEXT:

AN EXAMPLE OF THEORETICAL REFINEMENT

There has in fact already been much illuminating analysis of theoretical issues in FRBR, analysis which, as described above, surfaces rich connections with successful mature theories in other fields. We hope to survey this work, which has been valuable to us, at a later time. Here we discuss, from our own experience, just one specific case of interdisciplinary theoretical refinement.

In 2002 we noticed an interesting entity assignment puzzle: is an XML Document a FRBR manifestation or a FRBR expression?[105] There were good arguments on both sides. We concluded that the assignment depended on context of use, but it was unclear how to either reconcile this context-dependency with the FRBR ER model or revise that model to accommodate it.

We soon realized that this problem was similar to a previous puzzle about the taxonomy of descriptive markup.[103] That puzzle had been resolved by using the notion of illocutionary force (similar to grammatical mood), from linguistics (pragmatics in particular) to clarify an ambiguity:

orthographically and semantically identical XML markup could vary in illocutionary force depending on context of use. The context-dependency of manifestation/expression assignments described above seemed to be a partial generalization of this observation about XML markup. Later another related puzzle was noticed: some XML markup seemed to simultaneously refer to both the textual string which it had as content and also to the referent of that string.[106] We found that this puzzle was also illuminated by a notion from pragmatics, this time presupposition (the distinction between what is asserted, and what is presumed as a semantic condition of an assertion).

But these insights from linguistics, even in combination with other insightful work on theoretical problems with FRBR[93][94] still did not help us reconcile context-dependency with the FRBR model.

The critical clue came from computer scientists doing “applied ontology”. Guarino and Welty have developed a method for evaluating modeling decisions which requires that properties designating entity types be rigid in a sense defined using contemporary symbolic modal logic, but meaning, roughly, that the property in question is had permanently and essentially, not contingently.[96][97][98] Properties that fail this test should not be considered types of entities, but other sorts of things, such as roles which entities enter into in particular circumstances.

Originally we were interested in how these ontology evaluation rules might help with the Bechamel XML semantics project.[104] However it soon occurred to us to apply them to FRBR, and when we did we noticed that manifestation and expression both seemed to fail the rigidity test. This suggests that manifestation and expression are not

types of entities, strictly speaking. And on further reflection they in fact do seem more like roles that some types of entities may have in particular circumstances.[108]

Additional corroboration of this reinterpretation then came from aesthetics and the philosophy of social science:

Levinson’s analysis of musical works as natural (though abstract) objects in specific social context,[101] and Searle’s theory of social objects as natural objects in some specific social context[109]. We found both of these influential views compelling and felt they supported our emerging sense that the entity types identified by FRBR were not true types, but rather roles that other entities had in particular social circumstances.[108]

In light of these converging accounts, from multiple disciplines, we felt confident in conjecturing that a more conceptually precise version of the FRBR model would have FRBR entities as roles, not types.[108] This leads naturally to a number of new questions, such as what are the (true) entity types that take on these different roles, and what are the specific features of some social contexts in virtue of which they initiate and sustain these roles. We suspect that answering such questions will have us, again, drawing on work by linguists, computer scientists, philosophers, and others, as well as librarians and information scientists.

63. CONCLUSION

The preceding case illustrates by example how the theoretical refinement of a middle-range theory, such a FRBR, not only reveals and improves that theory’s explanatory power, but often does so by integrating perspectives from several disciplines

64. ACKNOWLEDGMENTS

As usual we acknowledge the contributions of members of three GSLIS research groups: the Electronic Publishing Research Group, the Research Writers Group, and the Metadata Roundtable, as well as the GSLIS Center for Research in Informatics and Science and Scholarship (CIRSS), directed by Carole Palmer. Participating in our discussions of FRBR during this period were Yunseon Choi, Thomas Dousa, Ingbert Floyd, Jin Ha Lee, Pat Lawton, Karen Medina, Christopher Phillippe Sara Schmidt, Richard Urban, Xin Xiang, Karen Wickett, Oksana Zavalina, and other members of the FRBR community. We also thank Michael Sperberg-McQueen (W3C/MIT) and Claus Huitfeldt (Bergen) for important early criticisms. Finally we thank members of the Balisage markup community, and, especially, Ann Wrightson, for suggesting that the work of Guarino and Welty could help us with our formalization XML document semantics. The usual disclaimers apply.

65. REFERENCES

- [92] Carlyle, A. 2006. Understanding FRBR as a conceptual model: FRBR and the bibliographic universe. *Library Resources & Technical Services* 50, 4.

- [93] Doerr, M., Hunter, J., and Lagoze, C. (2003). Towards a core ontology for information integration. *Journal of Digital Information* 4, 1.
- [94] Doerr, M., LeBoeuf, P. (eds). (2007) FRBR: Object-oriented definition and mapping to FRBR-ER. Version 0.8. International Working Group on FRBR and CIDOC CRM Harmonisation.
- [95] Floyd, I., and Renear, A. H. (2007). What in the digital world is a FRBR item? Poster at The Annual Meeting of the American Society for Information Science, October.
- [96] Guarino, N. and Welty, C. A. (2000). A formal ontology of properties. In *Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management. Lecture Notes In Computer Science*. Springer-Verlag.
- [97] Guarino, N. and Welty, C. 2002. Evaluating ontological decisions with OntoClean. *Communications of the ACM* 45, 2.
- [98] Guarino, Nicola and Chris Welty. 2004. An Overview of OntoClean. In Steffen Staab and Rudi Studer, eds., *The Handbook on Ontologies*.
- [99] International Federation of Library Associations. 1998. *Functional Requirements for Bibliographic Records: Final Report*. UBCIM Publications-New Series. Vol. 19, München: K.G.Saur.
- [100] Joint Steering Committee for Revision of Anglo-American Cataloguing Rules. 2006. RDA: Resource Description and Access. Frequently Asked Questions. Retrieved on June 20, 2006 from: <http://www.collectionscanada.ca/jsc/rdafaq.html>
- [101] Levinson, J. (1980). What a musical work is. *The Journal of Philosophy* 77, 5-28.
- [102] Merton, R. K. 1968. On sociological theories of the middle range, in *Social Theory and Social Structure*. The Free Press.
- [103] Renear, Allen H. 2001. The descriptive/procedural distinction is flawed. *Markup Languages: Theory and Practice* 2, 4.
- [104] Renear, A., Dubin, D., and Sperberg-McQueen, C. M. 2002. Towards a semantics for XML markup. In *Proceedings of the 2002 ACM Symposium on Document Engineering* (McLean, Virginia, USA) ACM, New York, NY.
- [105] Renear, A. H., Phillippe, C., Lawton, P., and Dubin, D. 2003. An XML document corresponds to which FRBR Group 1 entity? In: *Proceedings of Extreme Markup Languages 2003*. Montreal, Canada, August.
- [106] Renear A. H., Lee J. H., Choi Y., Xiang X. 2005. Exhibition: a problem for conceptual modeling in the humanities. In: *ACH/ALLC 2005. Conference Abstracts (2005) 2nd Edition*. Victoria: Humanities Computing and Media Centre, University of Victoria. 176–179.
- [107] Renear, A. H., Choi, Y. 2006. Modeling our understanding, understanding our models — the case of inheritance in FRBR. In: *Proceedings of the 69th ASIS&T Annual Meeting*, Austin, Texas.
- [108] Renear, A. H. and Dubin, D. 2007. Three of the four FRBR Group 1 entity types are roles, not types. : *Proceedings of the 70th ASIS&T Annual Meeting*, Milwaukee WI.
- [109] Searle, J. R. 1995. *The Construction of Social Reality*. New York: The Free Press.
- [110] Sutton, R. I. and Staw, B. M. 1995. What theory is not. *Administrative Sciences Quarterly*, 40.
- [111] Svenonius, E. 2001. *The Intellectual Foundation of Information Organization*. MIT Press.
- [112] Tillett, B. B. 2004. What is FRBR? A Conceptual Model for the Bibliographic Universe. Washington: Library of Congress, Cataloging Distribution Service. Online: <http://www.loc.gov/cds/downloads/FRBR.PDF>

FULL PAPER SUBMISSION 28:

The Role of the Arts in an I-School Education

AUTHOR(S):

Heckman, Robert
Snyder, Jaime

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Introduction

In this essay we make the argument that experience and interaction with the arts – aesthetic experience – should play an important role in the education of information professionals. We argue that this interaction will provide a useful and necessary complement to the more familiar scientific method that currently informs I-School professional education. We discuss the principles inherent in an arts-based approach to learning in I-Schools, and briefly sketch an agenda for future research and pedagogical development.

Professional education in Information Schools is predominantly technical. Our students learn the technology skills required for effective librarianship, information systems development, and network administration, among others. But just as the arts are becoming increasingly experimental and digital, so also are the information professions. The I-professional of the future may have to do much more than write a memo, a report, or make a standup presentation in order to communicate effectively. As professional communication increasingly gravitates to "new media," skills for professionals may come to rely increasingly on what we refer to here as the aesthetic way of knowing.

We make our argument for two fundamental reasons. First, the arts offer a complementary way of knowing that is not only highly beneficial for in the technical professions, but may also be important for the development of a competitive future work force. Second, we argue that much of the work done by technical professionals has more in common than would first appear with the work done by creative and performing artists, and thus is amenable to the pedagogical techniques employed in those fields.

The need for cross-functional skills

In discussions on how to prepare college graduates for future jobs, the argument is often made that today's market requires graduates to possess "cross functional" skills beyond those related to a specific professional expertise (e.g. Business-Higher Education Forum (BHEF) ,1997). Such skills transcend professional job categories, and are often missing in college graduates. The cross-functional competencies frequently identified are: leadership, teamwork, problem solving, critical thinking, global consciousness, tolerance for ambiguity, and basic communication skills (listening, speaking, reading, and writing). Interaction with the arts and exposure to an aesthetic way of knowing can provide one alternative model for I-School professional education.

The cycle of professional work

Our second argument is that work processes of technical professionals have much in common with those of creative and performing artists. Generalizing across many technical professions (e.g. librarianship, information systems design, software engineering, medicine accounting, engineering, etc.) Figure 1 illustrates four generic dimensions of professional work: production, performance, interpretation, and critique.

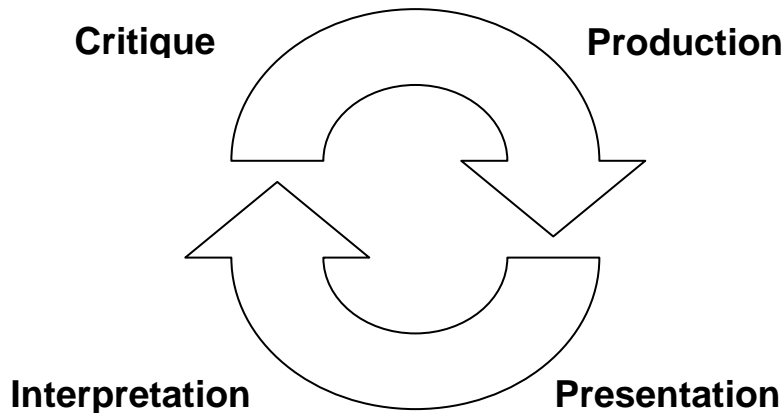


Figure 1. The Cycle of Professional Work

Production includes creation of professional artifacts. I-Professionals write programs, create designs, specifications, correspondence, proposals, requests for proposals, reports, and many other graphical and textual artifacts.

Performance in the I-Professions includes explicit role-based activities (e.g. interviewing to elicit requirements by reference librarians and system analysts, etc.) and also ubiquitous performances common to all professional work,(e.g. attending/conducting meetings, interviews, presentations, etc.).

Interpretation occurs whenever professionals are called upon to make sense of ambiguous situations, and to understand what is going on in times of uncertainty.

Critique involves the ongoing professional responsibility to evaluate the production and performance of others, and to make judgments about the professional value of such work.

It is easy to see the commonalities between professional and artistic production and performance. Beyond production and performance, however, there are less obvious similarities to professional work that draw our attention to the aesthetic realm in the areas of interpretation and critique.

While organizations strive for a controlled and predictive state of operation, they also demand innovation and invention as a means to increase resources and ultimately gain advantage. By recognizing that I-Professionals require the ability to use both

creative thinking and scientific rigor when confronting challenges, a technical education informed by arts-based practice can benefit both individuals and organizations.

Ways of knowing in a digital environment

Today's specialized knowledge and technology workers rely on a rational model of knowing based on scientific principles. This scientific approach has undeniably resulted in valuable technology that has allowed us to produce complex and sophisticated tools. However, the ability to use technology effectively does not necessarily evolve directly from the scientific, rational approach to knowing.

There is growing collection of literature that recognizes the importance of arming IS students with a broader range of skills in order to ensure success in a competitive market. Business management educators are looking to the liberal arts in order to bridge the gaps in IS education related to people-skills (Harrison & Akinc, 2000; Morris, Urbanski, & Fuller, 2005; Short & David J. Ketchen, 2005). The American Assembly of Collegiate Schools of Business (AACSB), has specifically advocated the integration of liberal arts education into the business curriculum (Harrison & Akinc, 2000).

Aspects of arts-based understanding

Engagement with the arts as a method for gaining understanding can help to address these gap. The artistic approach to knowing is characterized by intuitive understanding, aesthetic and emotional sensitivity, catharsis, metaphor and analogy. Artistic mastery is often indicated by a combination of creativity, eloquence and highly sophisticated craft.

Bruner identifies four aspects of the experience of art (Bruner, 1962):

- Connectedness – bridging the gaps between one's experiences in order to discover a cohesive view of the world
- Effort – departing from the habitual and literal ways of looking at things
- Conversion of impulse – a art can provide boundaries and forums for the exploration of seemingly uncontrollable instincts or impulses
- Generality – shock of recognition providing an expanded view of the known

Achieving understanding in these areas means comprehending at a deep level, in ways that lie outside the predictive and controlled realm of science. By providing opportunities to students that will allow them to experience these dimensions of knowledge, we can enable them to more successfully tackle socially complex and ambiguous problems. These are precisely the types of problems that they will likely encounter in the real world of organizational life.

Applying an arts-based approach to learning

Studio learning environments focus on idea generation, production and critique. Often there is no "right" answer. A student is expected to create, present and defend their work in front of the class. The class is then expected to provide thoughtful and constructive criticism to help improve the idea. Students take turns presenting what they have created and offering critique of the

work of others. Much of the learning is done “out in the open,” within a studio setting that provides a shared environment for mistakes, inventions and questions. It is common for problems posed to students to be ambiguous and equivocal. This setting provides students with opportunities to see multiple solutions, within a concentrated timeframe. Potential solutions can be evaluated in real time.

Teaching practices commonly used in studio and performance-based pedagogy can unlock this broader spectrum of experiences for I-School students, without compromising other aspects of their education. By tapping into these practices, I-School educators can introduce new methods for improving people skills and creating more flexible thinkers.

Returning to the cycle of professional work, at each stage, specific art-based principles can be introduced:

- Production
 - Exploit communal discoveries
 - Be aware of iterative and incremental progress
 - Utilize visibility
- Performance
 - Establish ownership
 - Exercise repetition
 - Observe results
- Interpretation
 - Consider group and individual perspectives
 - Encourage tolerance for ambiguity
 - Explore multiplicity of meanings
 - Determine goodness of fit
- Critique
 - Dialogue
 - Compare and contrast interpretations
 - Enable change and adaptation

Conclusion

The arts are moving in new directions, and the professions are moving into unknown territory as well. As the experimental arts evolve to include digital media, computer animation, computer music, sound art, and many novel combinations (Monaghan, 2007), so also are the information professions evolving to include unexpected combinations.

65.1.1 In this paper, we build on these ideas to first explore the theoretical foundations of an I-School pedagogy that integrates interaction and experience with the arts. We then describe pedagogical initiatives in a variety of technical fields that provide insight and guidance to those seeking to adapt arts-based techniques into I-School curricula. Finally, we sketch the outlines of an agenda for research-in-practice that will inform those who wish to systematically investigate arts-based techniques in order to contribute to a growing body of knowledge on the role of the arts in an I-School education.

References

- Bruner, J. (1962). *On Knowing: Essays for the left hand*. Cambridge, UK: Beklnap Press of Harvard University Press.
- Harrison, J. K., & Akinc, H. (2000). Lessons in Leadership from the Arts and Literature: A Liberal Arts Approach to Management Education through Fifth Discipline Learning. *Journal of Management Education*, 24(3), 391-413.

- Monaghan, P. (2007) Where High Tech Meets High Concept. Chronicle of Higher education February 23, 2007.
- Morris, J. A., Urbanski, J., & Fuller, J. (2005). Using Poetry and the Visual Arts to Develop Emotional Intelligence. Journal of Management Education, 29(6), 888-904.
- Short, J. C., & David J. Ketchen, J. (2005). Teaching Timeless Truths Through Classic Literature: Aesop's Fables and Strategic Management. Journal of Management Education, 29(6), 816-832.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

(e.g. Business-Higher Education Forum (BHEF), 1997). Such skills transcend professional job categories, and are often missing in college graduates. The cross-functional competencies frequently identified are: leadership, teamwork, problem solving, critical thinking, global consciousness, tolerance for ambiguity, and basic communication skills (listening, speaking, reading, and writing). The need for these skills is exacerbated by the fact that the US Department of Education is forecasting a shortage of college-educated workers in United States of more than 12 million by the year 2020. The nation is on the verge of a critical shortage of workers with college-level skills. The report calls for the United States to create 'a nation of learners'.

The report argues that the United States' competitive position in the world depends on its ability to educate a sufficient number of workers who are able to interpret complex and ambiguous situations, solve problems, communicate, and lead. To accomplish this, the nation needs major investments and major changes in the models of higher education that are used to prepare students for the job market.

We propose that an arts based approach to professional education can provide one such alternative model. Engagement with the arts as a method for gaining understanding can help to address this gap. Bruner [1] identifies four aspects of the experience of art that hint at the potential benefits of this approach:

- **Connectedness** – interaction with the arts can generate a synthetic response that helps one bridge the gaps between seemingly disconnected experiences in order to discover a cohesive view of the world.
- **Effort** – interaction with the arts requires work, work that often demands departing from habitual and literal ways of looking at things.
- **Conversion of impulse** – art can provide boundaries and forums for the experience of seemingly uncontrollable instincts or impulses, thus allowing for their exploration and understanding in ways that would be otherwise proscribed or inhibited.
- **Generality** – interaction with the arts often creates a shock of recognition that provides an expanded view of the known

These forms of understanding involve comprehending at a deep level, in ways that lie outside the linear, analytic mode of scientific method. By providing opportunities to students that will allow them to experience these dimensions of knowledge, we can enable them to more successfully tackle socially complex and ambiguous problems. These are precisely the types of problems that they will likely encounter in the real world of organizational life.

3. THE ARTS AND INTERDISCIPLINARY COLLABORATION

Our second argument stems directly from increasingly frequent opportunities we see at the iSchool at Syracuse University. Occasions for interdisciplinary collaborations within the school, within the university and across the wider academic community, as well as potential partnerships with industry, are springing up with great regularity. These opportunities cast a spotlight on the need to provide our students with a diverse set of problem-solving tools to help them build a strong and flexible set of critical thinking skills that will enable them to take part in these

innovative initiatives. Many of these opportunities involve working with colleagues in disciplines closely aligned with the information field, such as computer science and business. However, many are bringing us in contact with partners from the humanities. Additionally, we have observed that many of the work environments our students are entering upon leaving our school are increasingly driven by diverse teams that include members from a range of technology, design and marketing fields. The longevity of this trend is supported by Grant's [2] argument that organizations exist to provide integrating routines that allow individuals with diverse specialized knowledge to collaborate effectively.

Learning to participate in successful collaborations involves gaining so-called "people skills," including developing methods for resolving conflict and acquiring the ability to articulate and defend one's own ideas to people not necessarily sharing the same background or vocabulary. Exposure to the arts, as a complement to technical training, can give our students great advantage in these areas. Education in the arts, including literature, fine art and music, often focuses on 1) developing tools of empathy, 2) increasing awareness of the diversity of the human experience, 3) honing methods for reading subtext and nuanced meaning, and 4) cultivating an ability to interpret ambiguity from a number of vantage points.

All of these skills are desirable, if not in fact necessary, in collaborative work groups. Specifically, as we will discuss further, exposure to a studio art working and learning environment provides students with the opportunity to explore diverse expressions and creative solutions in a group setting. When these situations bear the most fruitful results, students struggle through their own attempt to express themselves while simultaneously interpreting and evaluating the work of their peers. The interdependence of studio cohort members can become one of the most generative aspects of the studio art experience. Rich stimuli in the form of artistic content can provoke discussions engineered to teach intellectual skills and studio-based learning can provide an opportunity to try out these new intellectual tools in a hands-on communal environment. The marriage of theory and practice that marks much arts-based learning has the potential to enhance the ability our students to work successfully in collaborative situations.

We reinforce this argument by presenting two concrete examples of opportunities for collaboration that would be greatly enhanced by providing participating students with exposure to the types of skills gained through experiences with the arts.

Colleagues at the iSchool at Syracuse have been advocating the creation of a profession devoted specifically to the support role that information professionals play in enabling people to make the most of new technologies. Referred to as a "cyberinfrastructure facilitator," one of the main responsibilities of this position is to advocate for and enable the use of cyber-based systems, tools, and services within any context, be it financial services, education, the arts or sciences. Obviously a large part of this vision is ensuring that the information professional has adequate communication skills and a full understanding of the needs and culture of the community that he or she is serving. We believe that arts-based experiences in the classroom can provide students with a forum to hone these skills and learn how to process diverse perspectives.

Another good example of the type of multidisciplinary project that would benefit from students well-versed in arts-informed learning is an initiative led by Michael McAllister, Director of Innovation and Design at the School of Art and Design within the College of Visual and Performing Arts (VPA) at Syracuse. He is advocating for the creation of a multimedia design center on campus that will enable students to get hands-on experience solving real design problems within the Syracuse community. In response to specific needs within the local community, such as education, outreach to underrepresented groups within the city, and small business/entrepreneurial startup programs, students will propose, develop and, importantly, implement large-scale multimedia projects. The program intends to bring together VPA students with those from the iSchool, as well as business, media and public policy areas. By partnering with VPA, we would provide our students with the chance to propose, facilitate and implement technology-driven solutions to an array of real world information problems.

The contributions that our information students could make to these projects are significant, not to mention the value that they would get from participating. But in order to make the most of these experiences, we need to make sure that we are teaching them the skills necessary to communicate and function within these diverse teams. We should not neglect our responsibility to make sure our students are properly prepared to fully participate in innovative programs like this, both while they are here at the iSchool and after they graduate.

In fact, we argue that many of the skills that are needed to take full advantage of opportunities like these are very successfully taught by utilizing the studio-based approach (described below) that advocates peer interactions, critical thinking, thoughtful and thought-provoking experimentation and openness to alternative solutions. These are also the same skills that many of our students will need when they enter the work force and find themselves working in multidisciplinary teams tackling similar problems.

4. THE CYCLE OF PROFESSIONAL WORK

Our third argument is that the work processes of technical professionals have much in common with those of design professionals and creative and performing artists. As a result, pedagogical approaches that have been found effective in education for the arts can also have a role in the education of iProfessionals. And just as the arts are becoming increasingly experimental and digital, so also are the information professions. For example, the iProfessional of the future may have to do much more than write a memo, a report, or make a standup presentation in order to communicate effectively. As professional communication increasingly gravitates to "new media," skills for professionals may come to rely increasingly on what we refer to here as the artistic mode of knowing and thinking.

While organizations strive for a controlled and predictive state of operation, they also demand innovation and invention as a means to increase resources and ultimately gain advantage. By recognizing that iProfessionals require the ability to use both creative thinking and scientific rigor when confronting challenges, a technical education informed by arts-based practice can benefit both individuals and organizations.

Generalizing across many technical professions (e.g. librarianship, information systems design, software engineering, medicine accounting, engineering, etc.) Figure 1 illustrates four generic dimensions of professional work: production, performance, interpretation, and critique. Technical professionals may cycle iteratively through all four stages in projects where they have heavy involvement, or may participate in one or two stages where their collaborative contribution is more peripheral.

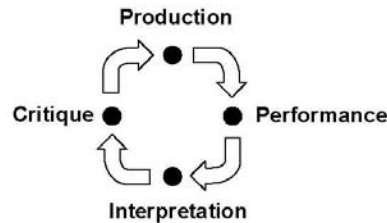


Figure 1. Cycle of professional work

Production includes creation of professional artifacts. Software engineers write programs. Systems analysts and engineers create designs and specifications. Accountants create financial statements. And professional workers of all types create professional correspondence, proposals, requests for proposals, reports, and many other graphical and textual artifacts.

Performance in the iProfessions includes explicit role-based activities (e.g. interviewing to elicit requirements by reference librarians and system analysts) and also ubiquitous performances common to all professional work, (e.g. attending/conducting meetings, interviews, presentations, etc.).

Interpretation occurs whenever professionals are called upon to make sense of ambiguous situations, and to understand what is going on in times of uncertainty. Each technical profession has its own set of analytical tools, based on the rational, scientific model to assist in this task. (e.g. decision trees to aid physicians in diagnosis, representational tools used by systems analysts to understand and model systems) But we argue that the rational, scientific ways of knowing that each profession brings to its field of focus are necessary, but not sufficient when interpretation involves complex human and organizational contexts.

Critique involves the ongoing professional responsibility to evaluate the production and performance of self and others, and to make judgments about the professional value of such work. Each profession has developed rational tools to assist in this process of evaluation (e.g. cost-benefit analysis, scoring models). Nevertheless we hear again and again that such evaluation tools are not always sufficient to fully evaluate products and performances in situations of high ambiguity and uncertainty, and where costs and benefits are strategic or less tangible.

It is perhaps easier to see the commonalities between professional and artistic production and performance due to the ubiquity of

professional artifacts and presentations. But the less obvious similarities in the areas of interpretation and critique may prove to be the more important because they involve skills central to professional judgment, skills that are more elusive and difficult to teach.

5. APPLYING AN ARTS-BASED APPROACH TO LEARNING IN I-SCHOOLS

At this point we introduce a practical example of how arts-related practices can be integrated into an existing iSchool curriculum. We divide methods for incorporating the arts into iSchool pedagogy into two categories: *content approaches* and *process approaches*. In future work we will discuss how to incorporate arts related content such as novels, poems, paintings or music as stimuli for discussion into professional courses. Examples of this type of augmentation to standard business school curricula can be found in the literature [10, 11, 13]. For this paper, however, we have selected a process-oriented strategy that introduces principles of studio art learning into a technical education classroom.

Teaching practices commonly used in studio and performance-based pedagogy can unlock a broader spectrum of experiences for iSchool students, without compromising other aspects of their education. In fact, educators in a range of professional and technical fields have sought to make use of studio-based techniques in the classroom [19-21]. By tapping into these practices, iSchool educators can discover new methods for improving people skills and creating more flexible thinkers.

Studio learning environments focus on idea generation, production and critique. It is common for problems assigned to students to be ambiguous and equivocal. A student is expected to create, present and defend their work in front of the class. The class is then expected to provide thoughtful and constructive criticism to help improve the idea. During discussion and evaluations, there is often there no "right" answer. Students take turns presenting what they have created and offering critique of the work of others. Much of the learning is done "out in the open," within a setting that provides a shared environment for mistakes, inventions and questions. This setting provides students with opportunities to see multiple solutions, within a concentrated timeframe. Potential solutions can be evaluated in real time.

Returning to the cycle of professional work, at each stage, the specific art-based principles can be introduced via a studio approach:

Production

- Exploit communal discoveries
- Be aware of iterative and incremental progress
- Utilize visibility

Performance

- Establish ownership
- Exercise repetition
- Observe results

Interpretation

- Consider group and individual perspectives
- Encourage tolerance for ambiguity
- Explore multiplicity of meanings

Critique

- Dialogue
- Compare and contrast interpretations
- Determine goodness of fit
- Enable change and adaptation

As an example of how a studio-based approach can be applied to an iSchool curriculum, we present an adaptation of a current undergraduate course regularly offered by our school: IST444 Information Reporting and Presentation. An informal survey of the curricula of other iSchools shows that many include a similar course, designed to introduce students to basic concepts related to successfully presenting information within a business context, in both oral and printed form [22]. Recognizing the sheer volume of information that the average professional handles in a typical work environment, this course provides students with tools, strategies and experience in making presentations with skill and competence. The course is designed to cover basic graphic design skills and software, and to provide regular opportunities for 3 to 5 minute oral presentations. Students delivery speeches, create information graphics, evaluate and discuss examples of printed and digital information.

With the permission of the IST444 professor of record, Susan Bonzi, we have reviewed the course syllabus and identified opportunities to align this class with the four stage cycle of professional work presented above. A studio-based iteration of the course, as described below, will be piloted in the fall of 2008 by the second author.

A typical instantiation of IST444 might follow a schedule similar to the one shown below (see Table 1), with individual and group presentations occurring throughout the semester. Notably the course is designed to introduce specific key concepts and provide introductory exposure to common graphics software. As such, the course takes a linear approaching, tackling each topic in sequence. The cumulative effect for students is exposure to all of the objective skills and techniques related to information presentation and reporting.

The IST444 course is well suited to incorporating a studio-based approach. Currently, students work in small groups to create profiles for fictional businesses and each group is expected to generate materials for their company during the course of the semester. Instruction in public speaking is an important aspect of the course and these mock businesses provide content for a series of oral presentations that emulate real-world business situations. Concurrently, graphic design issues related to creating business documents such as resumes, letterhead and business cards are discussed. The class is structured to allow significant time for individual and group oral presentations while concurrently introducing information presentation guidelines and basic instruction in software such as Adobe Creative Suite and Macromedia Flash.

Table 1. Typical class structure for IST444 Information Presentation and Reporting

Weekly topics	Phase of professional work cycle addressed
Individual and Group Presentations (take place throughout the semester)	Performance, Critique
Graphic Basics	Production
Type	Production
Logos	Production
Types of speeches	Performance
Constructing a speech	Performance
Presenting a speech	Performance
Photoshop Basics	Production
Advanced Photoshop	Production
Basic Page design	Production
In Design basics	Production
Advanced inDesign	Production
Letter head, business cards	Production
Brochures	Production
Newsletters	Production
Color, Paper	Production
Resumes	Production
Charts and graphs	Production
Info graphics	Production
Concept mapping	Production
Web Page design	Production
Flash	Production

Our goal in redesigning the course offerings is to generate a more cyclical plan that iteratively guides students through the four stages of professional work:

1. *Production*— Students generate ideas and objects.
2. *Performance*— Students present those ideas to their peers in a variety of formats
3. *Interpretation*— Students analyze peer presentations in a formal setting
4. *Critique*— Students evaluate the success of the presentations based on collective analysis and on their own experiences and opinions

Because the course already includes a degree of Production, Performance and Critique (see designations, Table 1), our focus is on introducing elements related to Interpretation, as well as restructuring course content to reinforce the creative cycle that underlies much studio-based course work.

Our alternate plan for the course, based on a studio approach, covers the same material over the course of the same amount of time, and exploits the power of repetition and iteration that is common to studio courses. Instead of the linear format shown above, the course could be organized as a series of iterative cycles, stepping students through the work cycle multiple times and while increasing exposure to skills of interpretation. Such a course might look something like this (Table 2):

Table 2. Example of IST444 Information Presentation and Reporting based on a studio approach

Topic	Phase of professional work cycle
Present yourself	
Create a profile for yourself	Production
Present a resume	Performance
Alter your resume for different jobs or situations	Interpretation
Evaluate peer resumes	Critique
Create a company profile	
Come up with an idea for a company	Production
Pitch your company to the group	Performance
Evaluate the strengths and weakness of your company	Interpretation
Group discussion	Critique
Create a brand identity for your company	
Branding and logo	Production
Run a focus group	Performance
Participate in a focus group	Interpretation
Group discussion	Critique
Present a product offered by your company	
Create an advertisement for a product	Production
Present the product to the group	Performance
Compare your product to competitors	Interpretation
Group discussion	Critique

While the specific skills and exposure to software are virtually the same as the more traditional linear approach, the iterative cycle allows students to become more familiar with the cycle itself and this integrated view of the process “in-action” will itself be a valuable tool.

We have heard from instructors who have tried this type of approach that the last two steps merge together and, in the worst cases, result in dead-end statements from students such as, “I like it.” One of the challenges of adopting this approach is to maintain a distinct separation between the analysis of a presentation, be it a speech or a resume, and the ultimate evaluation of its worth within a given context. For example, a group analysis of a specific logo may reveal that half the class deems it “childish” while the other half sees it as “whimsical.” Arriving at that conclusion would be achieved during the Interpretation phase of the cycle. However, when we move onto critique and we look at the company for which the logo was created, say a company that makes computer game consoles designed for adult users, we can then start to make value judgments. That is when we may conclude that a whimsical logo may be better suited than a childish one for specific reasons related to the target audience profile that was part of the initial company profile, and a discussion of how to optimize the graphic could follow.

The Role of the Arts in an iSchool Education

Robert Heckman
School of Information Studies
Syracuse University
320 Hinds Hall
Syracuse, NY 13244
315-443-1892
rheckman@syr.edu

Jaime Snyder
School of Information Studies
Syracuse University
337 Hinds Hall
Syracuse, NY 13244
315-751-3977
jasnyd01@syr.edu

ABSTRACT

Professional education in Information Schools is predominantly technical and based on a rational, scientific way of thinking about the world. In this essay we make the argument that experience and interaction with the arts – aesthetic experience – should play an important role in the education of information professionals. Our discussion of the role of the arts in an iSchool education is structured around three basic ideas. First, we argue that the arts offer a pathway into complementary mode of thinking and knowing that is not only highly beneficial for the technical professions, but may also be important for the development of a competitive future work force. We refer to this as an artistic mode of knowing, and we discuss both terminology and the characteristics in this section. Second, we build on the ideas of Grant and others that focus on the organizational and societal importance of collaboration among individuals with diverse specialized knowledge. We describe how the skills needed for successful interdisciplinary collaboration are increasingly drawing on both ways of knowing. Third, we argue that much of the work done by information professionals has more in common than would first appear with the work done by design professionals and creative and performing artists, and thus is amenable to the pedagogical techniques employed in those fields. We conclude by offering specific iSchool examples of arts-based pedagogy and include suggestions for incorporating these ideas into an existing curriculum.

Topics

Nature and scope of iSchools and iResearch

Keywords

arts, aesthetics, studio learning, information professionals, iSchools

1. INTRODUCTION

In this essay we make the argument that experience and interaction with the arts – aesthetic experience – should play an important role in the education of information professionals. We argue that this interaction will provide a useful and necessary complement to the more familiar rational, scientific model that currently informs iSchool professional education, and discuss the principles inherent in an arts-based approach to learning in iSchools.

Professional education in Information Schools is predominantly technical and based on a rational, scientific way of thinking about the world. Our students learn the skills required for effective librarianship, information systems development, and network administration, among others, and it is entirely appropriate that the primary focus of a professional education be on the technical knowledge and skill that professionals must master. But beyond technical competence, successful professionals also require the ability to interpret complex and ambiguous situations, interact with those expert in other forms of specialized knowledge, and constructively evaluate their own work and the work of others. These abilities call for an approach to knowing the world that complements the knowledge provided by a rational, scientific approach. This approach to knowing is characterized by intuitive understanding, emotional sensitivity, catharsis, metaphor and analogy. Bruner [1] proposes that the experience of art promotes an integrated, holistic view of phenomena that is a useful complement to analytic knowing.

We structure our discussion of the role of the arts in an iSchool education around three basic ideas. First, we argue that the arts offer a pathway into complementary mode of thinking and knowing that is not only highly beneficial for the technical professions, but may also be important for the development of a competitive future work force. We refer to this as an artistic mode of knowing, and we discuss terminology and describe this mode here. Second, we build on the ideas of Grant [2] and others that focus on the organizational and societal importance of collaboration among individuals with diverse specialized knowledge. We describe how the skills needed for successful interdisciplinary collaboration are increasingly drawn from diverse ways of knowing. Third, we argue that much of the work done by information professionals has more in common than would first appear with the work done by design professionals and creative and performing artists, and thus is amenable to the pedagogical techniques employed in those fields.

We explore these three assertions in detail in Sections 2, 3 and 4 of this paper. In Section 5 we offer a specific iSchool example of arts-based pedagogy and include suggestions for incorporating these ideas into an existing curriculum.

2. SCIENTIFIC AND ARTISTIC MODES OF KNOWING

For the past two centuries, specialized technical workers have relied on an approach to thinking and knowing derived from the scientific method of acquiring knowledge. This scientific approach has resulted in valuable technology that has allowed us to produce complex and sophisticated tools. It has created improvements in health, prosperity and standard of living that are unprecedented in human history. However, the unintended side effects of our technological revolution (e.g. potential for mass destruction, climate change, etc.) make it clear that the ability to use technology effectively does not necessarily evolve directly from the scientific approach to knowing that allows us to *make* technology. In this section, we briefly sketch the outlines of an approach to thinking, knowing, and understanding that we believe can serve as a highly useful complement to the scientific approach to gaining knowledge.

We label this approach the artistic mode of knowing in order to emphasize its distinctness from the scientific approach, but we recognize that the label has the potential to mislead. The two ways of knowing we consider do not represent a pure dichotomy, but rather lie on a continuum, with most thinking and knowing taking place somewhere in the middle, and employing both modes in varying degrees. Our objective is to promote a mode of thinking and knowing that operates in terms of a "both-and" opportunity rather than an "either-or" choice. Art historian James Elkins is promoting a similarly holistic approach in his book entitled *The Domain of Images* where he makes an argument for an inclusive study of images, both those deemed as art and those seen as more utilitarian (such as images used in science). In fact, he devotes several pages to discussing the inadequacy of many if not all art/non-art monikers, examining (and ultimately finding fault with) a range of terms for non-art images including: scientific, nonreligious, inexpressive, non- or extra-aesthetic, utilitarian, informational, nonrepresentational, schematic and notational [3]. Ultimately he finds a distinction based on practice to be the most useful for his purposes: art images are those generally studied by art historians. Non-art images are not. We have followed a similarly pragmatic approach to clarifying our distinction between artistic and scientific ways of knowing and thinking about problems. While we continue to search for the perfect terms for these two modes, we will let these terms stand for the moment and move on to describe in more detail what each represent in terms of this research.

We begin by presenting in bullet form a series of distinctions between the two modes of knowing and thinking:

- Scientific thinking discriminates between and analyzes details. Artistic thinking synthesizes details into a single holistic pattern.
- Scientific knowing focuses on category; artistic knowing focuses on relationship.
- Scientific knowing relies on analysis; artistic knowing relies on synthesis
- Scientific knowing segregates; artistic knowing integrates.

- Scientific knowing is sequential, literal, functional, textual, and analytic.
- Artistic knowing is simultaneous, metaphorical, aesthetic, contextual, and synthetic [4].

There is a persistent but disconnected stream of research and argument that supports the need for a mode of thinking and knowing to complement the scientific model that dominates professional higher education. For example, there is growing collection of literature that recognizes the importance of arming management students with a broader range of skills in order to ensure success in a competitive market [5-8]. Business school educators are looking to the liberal arts in order to bridge the gaps in Information System education related to people-skills [9-13]. The American Assembly of Collegiate Schools of Business (AACSB), has specifically advocated the integration of liberal arts education into the business curriculum [14].

Researchers have also begun to explore the aesthetic dimension of information systems design. For example, Tractinsky [15] writes of the aesthetics of the information system design object (by aesthetics he means the common every day meaning of the term: "an artistically beautiful or pleasing appearance"). He notes that research in information systems has almost completely ignored this notion of aesthetics in order to focus on robustness and functionality, and speculates that this might be because the origins of the computing disciplines emphasize the rational model of hard science, efficiency, and utility. As information systems become more interactive and place more emphasis on visual displays, he argues that visual aesthetics should receive more research and pedagogical attention for several reasons. First, as science and technology provide products that dramatically exceed the functional needs of most humans, the aesthetic dimension becomes more important. Second, aesthetic interpretations and evaluations are often hard to overcome because they precede and influence reflective and cognitive interpretations. Finally, he argues that while some aesthetic responses are innate and relatively invariant, most are learned and depend on culture, education, and other experiences. Because they are acquired, we should focus on their education.

Bennett et al. [16], writing from a business management perspective, note the research from cognitive psychology and organizational cognition that describes human limitations in the ability to understand and represent ambiguous concepts and situations. This focus on the bounded rationality of human problem solvers [17, 18] reminds us of the importance of being able to interpret and evaluate complex, ambiguous situations holistically, by using tools other than those that are rational, and analytical. Because human rationality is bounded or limited, an aesthetic way of approaching complex, ambiguous situations may provide a number of benefits, including the ability to usefully integrate information that at first appears to be fragmented, unrelated and thus ambiguous. Aesthetic understanding can help information professionals coherently interpret ambiguous situations by illuminating the correspondences and relationships within them.

In other discussions on how to prepare college graduates for future professional jobs, the argument has often been made that today's market requires graduates to possess "cross functional" skills beyond those related to a specific professional expertise

It is also important to stress that this approach does not need to assume that students are or will be graphic designers. While students' mechanical skills with software and visual design abilities may be limited, the stress here is on process and problem solving. Many of these students will at some point in their professional careers be interacting with experts in marketing and design. One of our goals is to provide them with the vocabulary and communication skills necessary to successfully collaborate with these other professionals, in order to make the most of the expertise around them. Additionally, and just as importantly, by providing information students with insight into a creative process, we will potentially enable them to participate more fully in these partnerships, making a wider impact by bringing their knowledge of information technology into new areas.

6. CONCLUSION

In this paper we have argued that experience and interaction with the arts should play an important role in the education of information professionals, and that this interaction will provide a useful and necessary complement to the more familiar rational, scientific model that currently informs iSchool professional education. We discussed the principles inherent in an arts-based approach to learning in iSchools. We showed how the work done by information professionals is similar to the work done by creative and performing artists, and those in the design professions. Finally, we offered an example of how an iSchool course is being redesigned using these ideas.

We hope this brief account serves as a stimulus for others to consider, debate, and modify these ideas. If an arts-based approach to technical and professional education is to bear fruit, much research and development must be done. In future work we will propose research and practical agendas to carry these ideas forward. The arts are moving in new directions, and information professions are moving into unknown territory as well. As the experimental arts evolve to include digital media, computer animation, computer music, sound art, and many novel combinations [23], so also are the information professions evolving to include unexpected collaborations with colleagues from other fields and opportunities to make contributions across many social and organizational contexts. We believe that the educational approach we have outlined will prepare our students to meet these challenges.

References

1. Bruner, J., *On Knowing: Essays for the left hand*. 1962, Cambridge, UK: Beknap Press of Harvard University Press.
2. Grant, R.M., *Toward a knowledge-based theory of the firm*. Strategic Management Journal, 1996. 17(10): p. 109-22.
3. Elkins, J., *The domain of images*. 1999: Cornell University Press London.
4. Pink, D.H., *A Whole New Mind*. Riverhead, New York, NY, 2005.
5. Adler, N.J., *The Arts & Leadership: Now That We Can Do Anything, What Will We do?* Academy of Management Learning & Education, 2006. 5(4): p. 486-499.
6. Bartunek, J.M. and I. Carboni, *A Time for Hope: A Response to Nancy Adler*. Academy of Management Learning & Education, 2006. 5(4): p. 500-504.
7. Tung, R.L., *Of Arts, Leadership, Management Education, and Management Research: A Commentary on Nancy Adler's "The Arts & Leadership: Now That We Can Do Anything, What Will We do?"* Academy of Management Learning & Education, 2006. 5(4): p. 505-511.
8. Denning, P.J. and R. Dunham, *The Core of the Third-Wave Professional*. Communications of the ACM, 2001. 44(11): p. 21-25.
9. Weick, C.W., *Out of Context: Using Metaphor to Encourage Creative Thinking*. Journal of Management Education, 2003. 27(3): p. 323-343.
10. Morris, J.A., J. Urbanski, and J. Fuller, *Using Poetry and the Visual Arts to Develop Emotional Intelligence*. Journal of Management Education, 2005. 29(6): p. 888-904.
11. Short, J.C. and J. David J. Ketchen, *Teaching Timeless Truths Through Classic Literature: Aesop's Fables and Strategic Management*. Journal of Management Education, 2005. 29(6): p. 816-832.
12. Dunne, D. and R. Martin, *Design Thinking and How It Will Change Management Education*. Academy of Management Learning & Education, 2006. 5(4): p. 512-523.
13. Weinrauch, J.D., *An Exploratory Use of Musical Metaphors to Enhance Student Learning*. Journal of Marketing Education, 2005. 27(2): p. 109-121.
14. Harrison, J.K. and H. Akinc, *Lessons in Leadership from the Arts and Literature: A Liberal Arts Approach to Management Education through Fifth Discipline Learning*. Journal of Management Education, 2000. 24(3): p. 391-413.
15. Tractinsky, N., *Toward the study of aesthetics in information technology*. Proceedings Twenty-Fifth International Conference on Information Systems, 2004: p. 771-780.
16. Bennett, R.H., *The Mind's Eye and the Practice of Management Envisioning the Ambiguous Robert H. Bennett, Walter J. Wheatley, E. Nick Maddox, William P. Anthony The Authors*. Management Decision, 1994. 32(2): p. 21-29.
17. March, J.G. and H.A. Simon, *Organizations*. 1958: John Wiley & Sons Inc.
18. Simon, H.A. and J.R. Hayes, *Understanding complex task instructions*. Cognition and instruction, 1976: p. 269-285.
19. Lister, B.C., *Interactive Distance Learning: The Virtual Studio Classroom*. in *Proceedings of the Third International IEEE Conference on Multimedia, Engineering and Education*. 1998. Hong Kong.
20. Lister, B.C., *Next Generation Studio: A New Model for Interactive Learning*. Syllabus Northeast, 2001.
21. Wilson, J.M. and W.C. Jennings, *Studio Courses: How Information Technology Is Changing the Way We Teach, On Campus and Off*. Proceedings of the IEEE, 2000. 88(1): p. 72-80.
22. Bonzi, S., *Coure syllabus for IST-444 Information Reporting and Presentation*. 2007, School of Information Studies, Syracuse University.
23. Monaghan, P., *Where High Tech Meets High Concept*. Chronicle of Higher Education, 2007. 53(25): p. 1.

FULL PAPER SUBMISSION 29:

Information Practices to Sustain Knowledge Building: The Case of the Virtual Math Teams Online Community

AUTHOR(S):

Sarmiento, Johann

Stahl, Gerry

ORGANIZATION(S):

Drexel University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

**Information Practices to Sustain Knowledge Building:
The Case of the Virtual Math Teams Online Community**
Paper submitted for the Third Annual iConference 2008

The sustained knowledge building of virtual groups and online communities requires that co-participants overcome a wide range of gaps in their interactions, especially in the context of long-term activity across multiple episodes and collectivities. Here we present an analysis of the collective information practices of virtual teams engaged in collaborative problem-solving as part of the Virtual Math Teams (VMT) online community. Our analysis aimed at understanding the information practices that teams employed to *bridged* the apparent discontinuity of their collaborative interactions (e.g. multiple collaborative sessions, teams, and problem tasks) and exploring the role that such *bridging activity* plays in their knowledge building over time.

Knowledge Building and Collective Information Practices

Knowledge Building has been defined as the creation, testing, and improvement of conceptual artifacts (Scardamalia & Bereiter, 2006). In this sense, *knowledge building* is primarily information activity (individual and collective) carried out through a set of practices aimed at developing and advancing a person or a group's understanding of a question, a problem, a decision, or any other knowledge gap. Knowledge building is at the core of many human contexts including, for example, educational settings (Scardamalia, 2002; Stahl, 2006a) digital libraries (e.g., Bearman, 2007; Fox & Urs, 2002) and virtual communities (e.g., Barab, Kling, & Gray, 2004; Ellis, Oldridge, & Vasconcelos, 2004; Renninger & Shumar, 2002). Online contexts, in particular, make knowledge-building more visible and, as such, represent very unique opportunities to advance our understanding of how individuals, small groups and the larger community build, evolve, and expand knowledge.

Naturally, knowledge building is a complex and challenging enterprise. Divergent perspectives often lead to problems of understanding; different work styles can result in interactional breakdowns, while the distributed nature of joint action over time usually leads to gaps of awareness and problems of coordination. Discontinuities emerging from long-term knowledge building have been studied from a number of different perspectives. The gaps that arise among events, perspectives, and participants have been an area of investigation in the study of individual and group creativity (e.g. Amabile, 1983; Sawyer, 2003) as well as in fields such as Small-group Research (Arrow et al., 2000; Bluedorn & Standifer, 2004), Computer-supported Cooperative Work (CSCW) and Knowledge Management (Greenberg & Roseman, 2003; Ishii et al., 1993). Despite their interest in this crucial topic, most studies have concentrated solely on characterizing the outcomes of groups and communities who successfully overcome discontinuities but few descriptions have been offered of the information practices that lead to such outcomes. Among these outcomes we can list the existence of "information bridgers" in group-to-group collaboration (Mark, Abrams, & Nassif, 2003), the use of boundary objects in interdisciplinary collaboration (Star, 1989), the emergence of "shifting epistemologies" (Bielaczyc & Blake, 2006), and the growing orientation to collective knowledge advancement in communities (Scardamalia, 2002). Our interest lies in characterizing how

specific information practices are used to overcome relevant gaps in collective knowledge building.

Case Study: The Math Forum and the Virtual Math Teams (VMT) online communities

The Math Forum (<http://mathforum.org>) is an online community, active since 1992. It promotes technology-mediated interactions among teachers of mathematics, students, mathematicians, staff members and others interested in learning, teaching and doing mathematics. As the Math Forum continues to evolve, the development of new interaction supports becomes essential for sustaining and enriching the mechanisms of community participation. As an example, the Virtual Math Teams (VMT) project investigates the innovative use of online collaborative environments to support effective mathematical work by small groups. In the VMT project, small groups of students come together to work through a special online environment that provides them with an array of tools to conduct their collaborative problem-solving activity, sustain it over time, and interact with other interested individuals and groups (Wessner *et al.*, 2006).

During the spring of 2005 and 2006 we conducted two case studies of VMT activity to explore issues of continuity and sustainability of collaborative knowledge building. In each case, five virtual teams were formed with about four non-located secondary-school students selected by volunteer teachers at different schools across the United States. The teams engaged in online math discussions for four hour-long sessions over a two-week period. They used the VMT virtual room environment (Wessner *et al.*, 2006) which combines a persistent chat tool with a shared whiteboard in addition to some other interactional supports. At the start, the teams were given a brief description of an open-ended mathematical situation and were encouraged to generate and pursue their own questions about it. Later on, the teams were given feedback on their prior work and the work of other teams and were encouraged to continue their work.

The goals of our analysis were to understand how teams of participants in the VMT online community managed the apparent discontinuity of their interactions (e.g. multiple collaborative sessions, teams and tasks), and to explore the relationship between such activity and the teams' knowledge building over time. We employed the approach of ethnomethodology (Garfinkel, 1967) to examine the sequences of events by using recordings and artifacts from each team sessions. As part of the phenomenological perspective, ethnomethodology is based on naturalistic inquiry to inductively and holistically understand human experience in context-specific settings (Patton, 1990). For our purposes, we examined each of the 37 sessions recorded, paying special attention to the sequential unfolding of the problem-solving episodes in which each team participated. Constant comparison through the entire dataset led to our refinement of the structural elements that define the information practices presented below.

Information Practices: Local, Longitudinal and Expansive Continuity

Our study of bridging practices revealed a series of information practices which allowed the teams to cross over the boundaries of time and link together different episodes of

collective action. The three most significant practices documented in our analysis are: *Framing Narratives of past activity as resources for new tasks*, *Collective remembering*, and *Projective Memory*.

By *framing narratives of past activity as resources for new tasks*, teams were able to construct and use “*reportables*” such as rules, procedures, or results that had been discovered in prior activities and construct new tasks for their current activity. In doing this, participants seemed to contribute to the creation of a team history constructed out of episodes of interaction and the relevant team members’ biographies. *Projective Memory*, the least common of the three practices documented, involved the planning and configuration of future work out of incomplete work (e.g. missing explanations), further challenges (e.g. more complex problems), or possible new tasks. Finally, through *Collective Remembering*, participants and non-participants of prior activities engaged in reconstructing past results by juxtaposing past and present resources and, in many cases, advancing the frontiers of prior results.

These information practices seemed to have one interesting structural feature in common. They were deployed by interweaving three central aspects of the groups’ interactions: *time and sequences of activity*, *participation frameworks*, and *knowledge artifacts*. The *temporal or sequential organization of experience* (i.e. what was done in a different episode of activity or at a different time, how does one action relates to something done before, etc.) provided the underlying structure for the teams to construct a sense of continuity that was “*local*” within an episode of interaction, “*longitudinal*” as it related to the history of the team, and “*expansive*” as it projected towards the past and future activities of other groups in the online community. By enacting specific *participation frameworks* (e.g. positioning participants as being or not involved in a specific past present or future activity, as those who can or should speak about a particular matter, etc.) teams managed their ongoing dynamic interactions and the creation and management of *knowledge artifacts* such as tasks and other problem-solving resources. These three elements were central to the information practices related to continuity.

In summary, bridging activity allowed the teams to cross over the boundaries of time, activities, collectivities, or perspectives. Bridging then seems to define a set of information practices through which participants deal with the discontinuities relevant to their joint activity. Another way to describe bridging might be to say that they are concerned with what Roschelle and Teasley (1995) described as the construction and maintenance of a “joint problem space.” Originally this space was theorized as a “shared knowledge structure” that integrated goals, descriptions of the current problem state, and the awareness of available problem solving actions. In our case, we have seen that this joint problem space is significantly distributed over time, across participants and artifacts, and also that specific information practices are related to its construction, maintenance and use.

The successful construction and maintenance of a joint problem space constitutes a central challenge of effective collaborative knowledge building (Roschelle & Teasley, 1995; Stahl, 2006b; Suthers, 2005). Several studies in the field of computer-supported

collaborative learning have shown that the interactional manner in which this intersubjective problem space is created and used determines the success of the collaborative experience (e.g., Barron, 2003; Chi, 2000; Dillenbourg *et al.*, 1995; Hausmann, Chi, & Roy, 2004; Koschmann *et al.*, 2005; Wegerif, 2006). Our present analysis has shown that this challenge escalates in contexts characterized by longitudinal activity across multiple collectivities but that specific information practices can be deployed for teams to manage this complexity. To establish continuity and sustainability, virtual teams and online communities “bridge” multiple elements of their interactions continuously a very consequential undertaking that lies at the core of effective collective knowledge building.

References

- Barab, S. A., Kling, R., & Gray, J. H. (Eds.). (2004). *Designing for virtual communities in the service of learning*. Cambridge, UK: Cambridge University Press.
- Barron, B. (2003). When smart groups fail. *Journal Of The Learning Sciences*, 12 (3), 307-359. Retrieved from <Go to ISI>://000183896600001
- Bearman, D. (2007). Digital libraries. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 223-272): American Society of Information Science and Technology (ASIS&T).
- Chi, M. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161-238). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1995). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). Oxford, UK: Elsevier.
- Ellis, D., Oldridge, R., & Vasconcelos, A. (2004). Community and virtual community. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 145-186): American Society of Information Science and Technology (ASIS&T).
- Fox, E. A., & Urs, S. R. (2002). Digital libraries. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 503-589): American Society of Information Science and Technology (ASIS&T).
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Hausmann, R., Chi, M., & Roy, M. (2004). *Learning from collaborative problem solving: An analysis of three hypothesized mechanisms*. Paper presented at the 26nd annual conference of the Cognitive Science society Proceedings pp. 547-552.
- Koschmann, T., Zemel, A., Conlee-Stevens, M., Young, N., Robbs, J., & Barnhart, A. (2005). How do people learn? In F. H. R. Bromme & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication*. Amsterdam: Kluwer Academic Press.

- Mark, G., Abrams, S., & Nassif, N. (2003). Group-to-group distance collaboration: Examining the 'space between'. In K. Kuutti, E. H. Karsten, G. Fitzpatrick, P. Dourish & K. Schmidt (Eds.), *Proceedings of the 8th european conference of computer-supported cooperative work (ecscw 2003)* (pp. 99 - 118). Helsinki, Finland.
- Renninger, K. A., & Shumar, W. (2002). *Building virtual communities*. Cambridge, UK: Cambridge University Press.
- Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69-197). Berlin, Germany: Springer Verlag.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*. Cambridge, UK: Cambridge University Press.
- Stahl, G. (2006a). *Group cognition: Computer support for collaborative knowledge building*. Cambridge, MA: MIT Press.
- Stahl, G. (2006b). Sustaining group cognition in a math chat environment. *Research and Practice in Technology Enhanced Learning*, 1 (2), 85-113.
- Suthers, D. (2005). *Technology affordances for intersubjective learning: A thematic agenda for cscl*. Paper presented at the international conference of Computer Support for Collaborative Learning (CSCL 2005), Taipei, Taiwan.
- Wegerif, R. (2006). A dialogic understanding of the relationship between cscl and teaching thinking skills. *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, 1 (1), 143-157.
- Wessner, M., Shumar, W., Stahl, G., Sarmiento, J., Muhlpfordt, M., & Weimar, S. (2006). *Designing an online service for a math community*. Paper presented at the International Conference of the Learning Sciences (ICLS 2006), Bloomington, Indiana.

FULL PAPER SUBMISSION 30:

I-School Curricula: How Wide? How Deep?

AUTHOR(S):

Thompson, Richard A.

ORGANIZATION(S):

University of Pittsburgh, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Personal Comment:

I recently stepped down after serving as Chair of Pitt's Telecom Program for 16 years. The "fit" of a Telecom Program inside SIS at Pitt has been an ongoing discussion for most of my tenure. Now, as SIS tries to conform to the evolving definition of what an "I-School" is, the discussion of this "fit" has taken on even greater relevance. In my discussions with Dean Larsen in the last several years, I tried to cast the problem to be part of a larger one, which is: just how "wide" and "deep" should an I-School's curriculum be? This talk/paper represents the state of this "larger" discussion, which Dean Larsen and I believe might be relevant to all I-Conference attendees.

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The various academic disciplines are shown to occupy the space of human knowledge. One objective of this talk/paper is to propose how the I-School disciplines fit among the neighboring disciplines, and which disciplines are close or distant. The various classic disciplines are seen to distribute differently over a "Level of Abstraction" dimension (six levels are defined) in which Philosophy is mostly abstract and Engineering is mostly concrete. Then, this talk/paper discusses the I-disciplines' abstraction-profile in significant detail. Curricular "depth" is defined as the number of layers of abstraction that should be included (I argue for five of the six).

It is observed that the I-disciplines have different discipline-neighbors at different levels of abstraction. For example, at the highest level of abstraction, the near neighbors are Philosophy (for Ethics) and Sociology (Maslow's needs) and, at the lowest level of abstraction, the near neighbors are Computer Science and Electrical Engineering (implementing Information systems). In between, our near neighbors are Psychology (cognition), Business, Law, Literature, Political Science, and others. The role of a "Professional School" is factored in. Then, the case is made that the I-School's "sub-disciplines" (six are defined) may have different abstraction-profiles and different neighbors.

Curricular "width" is defined as how close our curricula should get to those of our near neighbors. Finally, "holes" are identified in the spaces between the I-disciplines and their neighbors. These holes represent significant areas of human knowledge, which are typically excluded from the academic disciplines at most universities. Professionals who have filled these holes have done it, perhaps poorly, by their own cleverness. This talk/paper argues that the I-Schools should strive to fill those holes, but doing so may require even greater curricular width.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

I-School Curricula: How Wide? How Deep?

Dr. Richard A. Thompson
Telecommunications Program
School of Information Sciences
University of Pittsburgh, 15260
US (412) 624-9423

thompson@sis.pitt.edu

ABSTRACT

The emerging discipline we call Information Science is shown to fit among its neighboring disciplines. After proposing any academic discipline's six levels of abstraction, and using them to define curricular "depth," detailed investigation reveals that IS butts against different neighboring disciplines at different levels. Then, after proposing IS' six sub-disciplines, they are shown not only to have different distributions across their levels of abstraction, but also to butt up against different neighboring disciplines. An I-

School's role as a "professional school" is factored into this discussion.

Curricular "width" is defined as how close a curriculum should get to its near neighbors. Then, "holes" are identified in the spaces between the I-disciplines and their neighbors. These holes represent areas of human knowledge, which are typically excluded from all the academic disciplines at most universities. It is argued here that I-Schools should strive to fill those holes by expanding their curricular width.

Topic

Nature and scope of iSchools and iResearch

Keywords

Curriculum, Discipline, Information Science

66. INTRODUCTION

The paper is introduced on a personal note. I recently stepped down after serving for 16 years as the Chair of the Telecommunications Program at the University of Pittsburgh. One of my more important activities during these years has been to participate with the chairs of the nation's other "Telecom" programs in defining Telecom as an academic discipline. Similar to the consortium of "I-Schools," the nation's "T-Programs" have tried to organize themselves, most recently under the banner of the International Telecommunications Education and Research Association (ITERA). A corresponding paper [1], which asks the same questions of Telecom Programs that this paper asks of I-Schools, will be presented at the ITERA-08 Conference.

A related ongoing discussion during my tenure as Telecom Program Chair has been how my Telecom Program "fits" inside the School of Information Sciences at Pitt. Now, as SIS tries to conform itself to the evolving definition of what an "I-School" is, the discussion of this "fit" has taken on even greater relevance. In my discussions with Dean Larsen in the last several years, I have tried to cast the problem to be part of a larger one, which is: how "wide" and how "deep" should an I-School's curriculum be? This paper represents the state of this "larger" discussion, which Dean Larsen and I believe might be relevant to all the I-Schools.

So, this paper proposes a two-dimensional universe for academic disciplines, attempts to place IS in this universe, and discusses the "fit" and the "holes." The paper is not the "be-all and end-all" on this topic. Instead, it is intended to act as a framework that might guide the discussion which completes what this paper only begins. The paper asks more questions than it answers and, in fact, six questions are explicitly stated.

67. THE SPACE OF HUMAN KNOWLEDGE

Academic disciplines occupy a multi-dimensional space of human knowledge. This paper focuses on two of these dimensions, as shown in Figure 1 below.

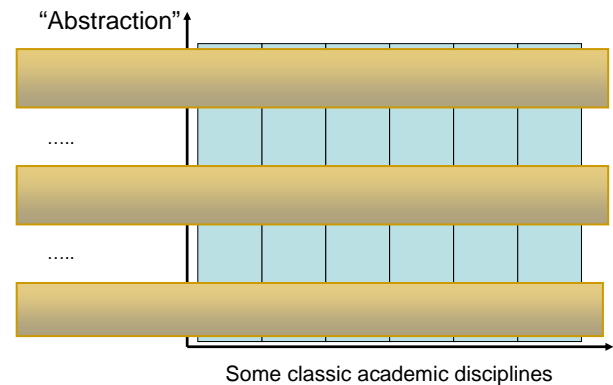


Figure 1. The Space of Human Knowledge

Two Dimensions

67.1 The Content Dimension, shown horizontally, is a spectrum that runs from the Fine Arts, through the Humanities and Social Sciences, to the Hard Sciences. The six columns represent only a few of the hundred or more recognized academic disciplines.

67.2 The Abstraction Dimension, shown vertically, is a spectrum of the level of detail and theory that is typically investigated in each respective discipline. While Figure 1 shows only three levels of abstraction, this paper is based on these six:

- Rationale – a discipline's broad justification.
- Framework – broad theory that establishes terminology in which to base generalized discussion and a structure for the discipline.
- Practice – the details that are needed by the people who will work in the field of endeavor that corresponds to the discipline.
- Optimization – detailed generalization and theory for the purpose of proving results, and determining various optimizations and best practices.
- Implementation – study of underlying infrastructure.
- Fundamentals – basic underlying principles.

For example, consider Education as an academic discipline. Its Rationale includes the discipline's rather self-evident purpose and motivation and its Framework includes issues like Curriculum design (this paper). Its Practice includes issues like classroom skill, boardmanship, enunciation, etc. and its Optimization is the body of knowledge, typically covered in a course on "Ed Psych," whose principles would presumably help us be better teachers and write better books. Education's Implementation level includes topics like knowing how to use Power-Point. The Education Discipline has few underlying Fundamentals like we find in Physics (Quantum Theory) or Biology (Genetics); one might be the elusive definition of learning.

The specification of these levels is certainly arguable; some might be combined and more might be defined. The discussion in this paper is based more on agreeing that academic disciplines have levels like these than on the details of what the levels are. However, the next sub-section provides a little justification.

Question 1: Are these the appropriate levels of abstraction?

2.2 Sophistication and Theory

Two other concepts related to abstraction are sophistication and theory. Sophistication relates to the maturity of the student, while abstraction relates to the applicability of the material. As shown in Figure 2, they are not unrelated. Mature scholars appreciate

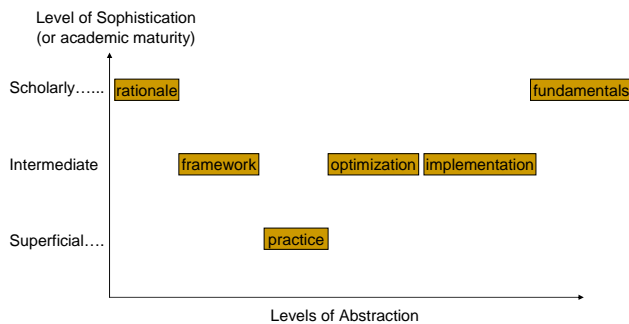


Figure 2. Sophistication v Abstraction

abstract issues like rationale and fundamentals, while immature students demand to know “what good is this?”

This paper separates theory into three layers, the even-numbered levels of abstraction in the list above: framework, optimization, and fundamentals. Level 2’s separation from Level 4 is based on an observation from Einstein in his 1933 essay, *On the Method of Theoretical Physics*:

“I want now to glance for a moment at the development of the theoretical method, and while doing so especially to observe the relation of pure theory to the totality of the data of experience. Here is the eternal antithesis of the two inseparable constituents of human knowledge, Experience and Reason, within the sphere of physics. Pure logical thinking can give us no knowledge whatsoever of the world of experience; all knowledge about reality begins with experience and terminates in it. Conclusions obtained by purely rational processes are, so far as Reality is concerned, entirely empty. But if experience is the beginning and end of all our knowledge about reality, what role is there left for reason in science? A complete system of theoretical physics consists of concepts, and basic laws to interrelate those concepts, and of consequences to be derived by logical deduction. Reason gives the structure to the system; the data of experience and their mutual relations are to correspond exactly to consequences in the theory.”

Paraphrasing Einstein, experience (Level 3) without underlying theory (Levels 2 and 4) is a disconnected set of

observations, and theory without observable verification gives no practical knowledge. It is remarkable that a theoretician, as Einstein was, would assert that theory is subservient to practice; and that its role is to help us understand practice and it provide a framework for organizing our observations. However, since not all theory helps to frame our disciplines, this paper separates theory into two levels at least.

Placing the IS Discipline

The various classic disciplines are distributed differently over the levels in the Abstraction Dimension. Consider three examples: Philosophy is mostly Rationale and the discipline concentrates at the upper levels; Engineering is mostly Practice and the various Engineering disciplines concentrate in the central levels; Mathematics, while shown as an upright rectangle in Figure 1, is the Foundation of the sciences and, as such, this discipline concentrates in the lower levels.

In this paper, “depth” is a relative term, defined in the Abstraction Dimension. Depth is not the absolute location of a discipline’s center of gravity along this dimension but is, instead, the number of levels that receive significant coverage in a given discipline. By this definition, Philosophy, Engineering, and Mathematics are not very deep disciplines. Examples of deep disciplines are Music and most of the Social Sciences. Physics, which fully engages all six levels of abstraction, is perhaps the deepest of all the academic disciplines.

The objective of this paper is to propose how our new, emerging discipline fits into the space of human knowledge with the other, more established, academic disciplines. Our broadly defined Information Science discipline must be made to fit in horizontally and vertically. The IS discipline’s place in the total space of human knowledge has “width” and “depth” as illustrated in Figure 3. We determine our depth by identifying how our discipline distributes itself across the vertical “abstraction” axis and identifying our content in each level. We determine our width by specifying the location of IS’ center of gravity on the horizontal “discipline” axis, identifying which disciplines are our neighbors, and identifying how closely our discipline butts up against these neighbors.

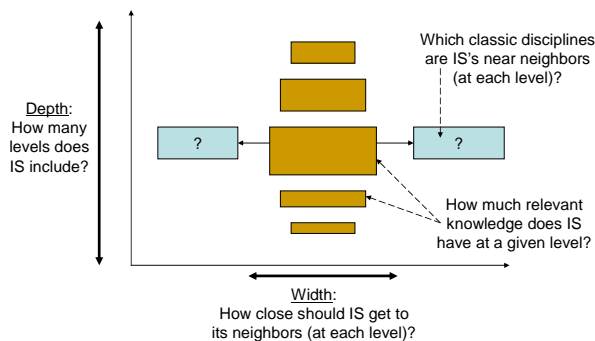


Figure 3. Depicting the Four Questions

Unfortunately, it's not so simple. Perhaps, some day it will be, when our discipline becomes more mature. But, now, these tasks are complicated because (1) the disciplines we identify as our neighbors are different at different levels of abstraction and (2) the distribution across the "abstraction" dimension is different in our different sub-disciplines. Section 3 begins to discuss the issues, and begs for much more follow-up.

68. HOW DEEP AND HOW WIDE?

This section uses the following procedure:

1. Identify IS' significant sub-disciplines.
2. Specify the IS discipline's levels of abstraction and how much content (depicted as the level's thickness) we have in each level.
3. Identify the neighboring disciplines for each level and recommend the width of each level.

The IS Sub-Disciplines(?)

Like everything else in this paper, this topic requires a lot of discussion and the proposal here may even be a little contentious. But, for our purposes, this paper assumes that Information Science is a "super-discipline" (see Section 4.2), comprised of the following six "sub-disciplines:"

- Librarianship
- Information Organization
- Information Management
- Information Systems
- Telecommunications
- Information Security

We make no attempt here to define or differentiate these; that is an important but separable, discussion. The point here is not that we agree that Information Science has these exact six sub-disciplines, but simply that we acknowledge that IS has several sub-disciplines, and they are something like these six.

Question 2: Is this the complete list of IS' sub-disciplines?

Figure 4's vertical axis shows the six proposed levels of abstraction. However, Figure 4's horizontal axis is a microscopic view of the disciplines of human knowledge,

illustrating the six sub-disciplines of our Information Science super-discipline. The figure serves to frame the discussion that follows. We will not discuss all 36 cells now; but eventually we must.

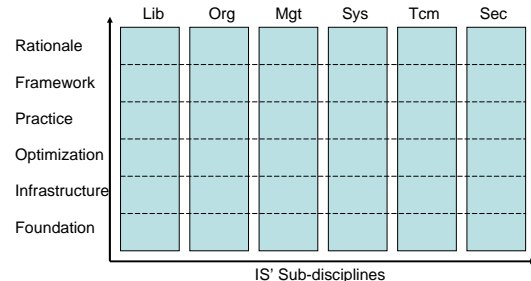


Figure 4. Framing the Content Discussion

3.2 How Deep?

The previous section described and defined six levels of abstraction. So, first, we examine how the IS super-discipline might distribute across these six levels. While we're at it, let's also discuss how much relevant knowledge (as depicted by the thickness of each horizontal band in Figure 3) we have at each level.

Rationale – Our Librarianship students typically understand our justification better than those in IS' other sub-disciplines. We don't have very much here, but all of our students should understand enough Ethics to resonate with the "Digital Divide" and they should understand how information fits into Maslow's hierarchy of human needs.

Framework – Much of the framework that organizes our discipline is found in a course commonly called "Information Retrieval." We need to develop this framework better, and ensure it's exposure to students in all our sub-disciplines, but we have a good start.

Practice – Each of the sub-disciplines described above focuses on the practical aspects of its corresponding profession. Since the I-Schools identify themselves as "Professional Schools," this is appropriate and it is our strength. IS' various sub-disciplines are appropriately separated in this level.

Optimization – Our discipline has little to offer at this level, but it's included in this list because Information Scientists must develop this if we are going to be taken seriously by the other disciplines. While we can optimize some things in Telecom, the other sub-disciplines must develop theory by which to compare two Information Systems and determine which one is better (whatever that means); or IS will forever be denigrated by our sister disciplines.

Implementation – Our Information Systems are built out of programs and electronic hardware. We have little to offer that is exclusively our own; but that's changing as the WWW, HTML, hypertext, and other implementation details are

perceived more-and-more as the purview of IS; and not CS. Our various sub-disciplines cover this material differently, if at all, depending on entrance requirements and students' skills.

Fundamentals – Our only true fundamental concept is Shannon's definition of Information. But, very few of us teach it and few of our students have the math skills to understand it, nor the sophistication to appreciate it. So, the amount of IS at this lowest level is slim-to-none.

Question 3: Are there other contributions in Levels 1, 2, 4, 5, and 6?

Since most of IS' content lies in Layers 2 and 3, the IS discipline is quite thin and will remain so unless we develop levels 1, 4, and 5 and try to teach level 6.

68.1 3.3 How Wide?

Now, let's examine where the various levels of our various sub-disciplines fit along the horizontal axis of human knowledge. We'll do this by identifying the neighboring disciplines at each level. This discussion is very incomplete; it presents only a small sample from the 36 cells defined in Figure 4.

Rationale – Our neighboring disciplines are Philosophy (when we discuss Ethics) and Sociology (when we discuss Maslow's needs). What we typically present is mostly borrowed from these neighbors, but we should develop it further and we should consider presenting this material to students in all our sub-disciplines.

Framework – Illustrating the diversity of our discipline, this level's many neighbors include: Psychology (for Cognition), Business (for managing information organizations and projects and for understanding applications in all organizations), Law and Political Science (for policy, intellectual property law, and e-government), Computer Science (for databases and data structures), and Literature (important in Librarianship).

Practice – While our Information Systems sub-discipline used to butt up against Business' MIS sub-discipline and our Information Systems and Information Security sub-disciplines butt up against Computer Science, we have succeeded recently at carving out more of our own space at this level.

Optimization – Most of what we do at this level is borrowed from Computer Science; although Telecom also borrows heavily from Electrical Engineering for optimizable concepts like bandwidth and signal-to-noise ratio. Information Science is sadly lacking at this level and, as stated earlier, Information Scientists must develop this level if we want to be taken seriously.

Implementation – The neighboring disciplines are Computer Science (for implementing programs) and Electrical Engineering (for implementing information systems).

Fundamentals – Shannon's definition of Information is borrowed from Electrical Engineering, but we should cast it in a more general way. Since it is our only fundamental, we should consider teaching it in all our sub-disciplines, even if it

means sufficiently developing our students' math skills so they can understand it.

While the discussion above identifies IS' many neighboring disciplines, it does not discuss the boundary between IS and its neighbors. That discussion, while it logically follows directly, is so relevant to IS curriculum design that this paper separates it into a separate, and concluding, Section 4.

Question 4: Not so much a question but, we must fill in details from all 36 cells in Figure 4.

3.4 Variations

Figure 5 below summarizes much of this section's discussion. Again, it's not meant to be the final word on this topic, but it attempts to illustrate the profiles of two of our sub-disciplines and how we might structure the IS super-discipline if we offered it as three different academic degrees.

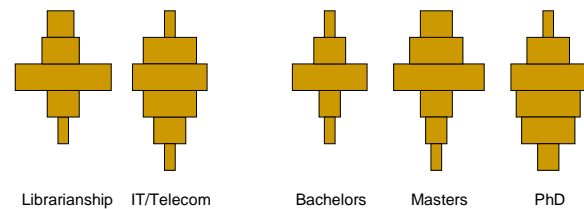


Figure 5. Variations by Sub-Discipline and Degree

69. WHERE DO WE GO FROM HERE?

Filling the Holes

Having described our near neighbors and the body of knowledge we have in common with them, let's discuss now how closely we should butt up against these neighbors. At each level of abstraction, we observe gaps or "holes" in the spaces between the I-discipline and its neighbors. These holes represent areas of human knowledge that are typically excluded from the academic disciplines at most universities. Many professionals, who graduated from our programs, have filled these holes themselves, perhaps poorly, by their own cleverness. Perhaps we could do a better job of preparing them for their professions by filling these holes for them while they are our students.

For example, those Telecom Program Chairs, who are trying to unify the nation's Telecom curriculum, struggle with what we call the Physical Layer. How well should a Telecom Professional understand Physical-Layer concepts like bandwidth, noise, filters, reflections, attenuation, and other forms of signal distortion? How well should they understand how signals propagate through wires, fibers, and free-space?

Many of us reside in universities that also have Electrical Engineering departments, where this material is taught,

but in a general way that also applies to electrical circuits, power systems, control systems, and other sub-disciplines of EE. Furthermore, when the EE's teach this material, they do it in a mathematical presentation that many Telecom students struggle with (in fact, the EET's split off from the EEs over this issue). Telecom students don't need that generality or mathematical presentation and EE students don't need the specific details in the Telecom application? So, there is no perfect course, the "hole" in Figure 6, unless we in Telecom teach it ourselves.

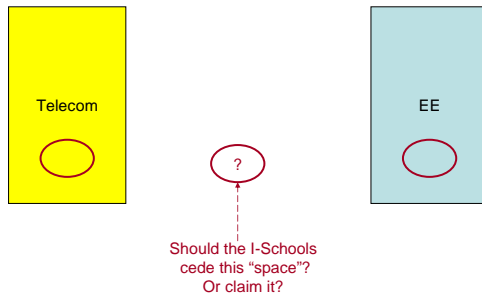


Figure 6. Example of a "Hole"

There are many other examples; only a few are listed below:

- Material on Information Ethics and on the human need for information could be part of the IS super-discipline; not left for one of its sub-disciplines and not left for the Philosophy or Sociology disciplines to offer.
- Many universities don't offer a course on human cognition, in IS nor in Psych. Our students need to know how the "human component" performs as part of an Information System. And, they would be better served by an IS approach to cognition.
- While the Business Schools at our respective universities might offer courses on MIS, Organizational Behavior (including the impact of technology on organizations) and Project Management, these are typically offered too generally. Our students need to understand the impact of information, and not all technologies, on organizations and they need to learn how to manage projects in their sub-disciplines.
- We have similar issues with the Law School and Political Science Department for course material on policy, intellectual property law, and e-government.
- Our problem with the Computer Science Department is even worse because the typical CS course on databases and data structures usually entails far more programming assignments than our students need.

- This list isn't complete without saying one more time that Information Scientists must develop a theory that supports Information Systems optimization. And, we should consider teaching Shannon's definition of Information in all our sub-disciplines, even if it means sufficiently developing our students' math skills so they can understand it.

This paper argues that the I-Schools should strive to fill these holes. But, there are two (at least) problems:

- Most IS faculty are very eclectic, but offering this list of courses will require even greater diversity and breadth.
- We get into the very difficult area of university economics and the pressure most of our universities are putting on us to "joint-list" more courses.

Question 5: Again not really a question, but we should compile a more complete list of the other holes we need to consider filling.

Defining the Super-Discipline

Our overall super-discipline should be greater than the intersection of the academic content covered by these six sub-disciplines, but small than their union. It should be some reasonable selection of the best parts from these six sub-disciplines. Then each of these sub-disciplines would be a specialization within this super-discipline.

When we look at the other schools and colleges at our respective universities, we see several models for how super-disciplines and sub-disciplines inter-operate. We need to consider which of these models is appropriate for us.

- In Arts and Sciences and in Engineering, about half the courses any student takes are common across all departments and about half are specific to a department.
- In most of the constituent departments, about three-fourths of what a student takes is common, and students may specialize in their final one or two semesters.
- In other schools, like Business, the curriculum is more isolated from the other disciplines. They may object to this characterization but, for example, while Math, Science, and Engineering students all take the same Calculus course, Business students have their own.

The I-Schools can adopt one of these models. This isn't much of an issue for an I-School that offers only one or two of the six identified sub-disciplines. But, if an I-School claims to cover the entire spectrum, or wishes to evolve in that direction, it might organize its curriculum as a common super-discipline with specializations, or as separate sub-disciplines with a small common core, or as something in between.

In SIS at Pitt, we offer these six sub-disciplines, but they have been “silo-ed” for too long. We are developing a com-mon introductory course, to be offered to all students next Fall Semester, and we are considering a common capstone course.

Question 6: What is “the right architecture” for the Information Science super-discipline?

70. REFERENCES

- [113] Thompson, R. 2008. The Telecommunications Curriculum – How Wide? How Deep? In the Proceedings of the ITERA-08 Conference (Louisville, KY, March 27-29, 2008). To be published.
<http://www.itera.org>.
- [114] Einstein, A. 1933. On the Method of Theoretical Physics. From Madden, E. H. (ed.). 1960. The Structure of Scientific Thought. Houghton-Mifflin, 1960 (essay originally published by Oxford University Press in 1933).

FULL PAPER SUBMISSION 31:

Culture At the Interface: Collaborative Design and Information Management with Indigenous Communities

AUTHOR(S):

Christen, Kimberly
Dietrich, Craig
Srinivasan, Ramesh

ORGANIZATION(S):

Washington State University
University of Iowa
University of California, Los Angeles

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Digital technologies provide the platform for both preserving and producing cultural knowledge and materials. Indigenous and other subaltern cultural perspectives on access, authority, and accountability in relation to cultural materials and knowledge are often marginalized in dominant narratives about the possibilities of digital technologies. Mainstream notions of “search” and “open access”—guided by Google’s mantra of providing universal access to everything—often leave out, or too easily blur, the lines between the multiple forms of technological translations and “other” knowledge/property systems. While digital technologies have much to offer to museums and other archival institutions and their many publics, their use needs to call in to question the viability of universal logics of search and access.

Since the mid 1990s, museums, archives, libraries and especially land grant universities in the United States have recognized the need to direct their energies towards outreach to the indigenous communities whose materials they hold. Many archives have signed Memorandums of Understanding (MOUs) with indigenous communities promising access and support in retrieving materials. Digital technologies have allowed archives to produce online databases and expand their physical catalogs with digital repositories. At the same time, archival mandates to preserve information intersect with the capabilities of digital technologies to distribute information over networks. In the context of “information freedom” rhetoric and appeals for “access to knowledge” archives have found themselves at the center of information management debates and strategies for new models for information indexing and distribution.

Museums and archives, now enlivened by the possibilities afforded by digital technologies, are at a crossroads in dealing with emergent information management systems and the cultural, ethical, and legal commitments they have to their multiple stakeholders. Making their materials “open” on the one hand, and working with divergent protocols for access and distribution on the other, In addition, physical distances, educational and linguistic barriers, and high levels of poverty have all made traditional archives unapproachable places for indigenous communities whose cultural materials and institutional histories are preserved within them. However, in the last ten years there has been a groundswell of activity lead by local indigenous communities in conversation with museums, archives, and technologists to incorporate indigenous cultural protocols into the parameters of archival sensibilities, design, and infrastructure. Projects around the world are seeking to integrate indigenous protocols into their informatics.

The papers in this panel focus on collaborative efforts between museums, archives and local indigenous communities in Australia and Native North America to create digital archival systems informed by indigenous practices and protocols. The panel members come from diverse academic and technological backgrounds representing the Fine Arts, Humanities and Social Sciences. Their work is informed by interdisciplinary concerns shaping scholars' understandings of the practical uses of digital technologies for underrepresented groups and the collaborative ways in which technology can be leveraged in service of cultural differences.

The first paper (co –presented by the principal investigator and software designer) based on ten years of collaborative work with Warumungu community members in Australia’s Northern Territory, examines the production of an indigenous community digital archive and its internal search functionality. The Mukurtu Wumpurrarni-kari archive, installed in the community in August 2007, leverages Warumungu protocols for the reproduction, distribution, and access to cultural materials and their associated knowledge within the design and structure of the archive. A series of restrictions and “sharing protocols”—designated by the Warumungu community—drives the searching, sorting and sharing of all materials in the system. The archive, thus, challenges dominant narratives of openness and accessibility in relation to intangible cultural materials and acts as a practical solution to issues of preservation, access, and distribution that face indigenous people globally.

The second paper describes a NSF-funded initiative which involves a direct partnership with the Zuni nation of New Mexico and the Cambridge (UK) Digital Museum Initiative. This partnership involves the study of how digital objects of excavated cultural materials can become the bases for indigenous engagement, and stimulate the release of indigenous and cross-cultural voices into the larger world of how digital systems are developed. It questions the assumptions that all “information is free” and should be widely disseminated without a sense of bounds, authorship, or culturally-specific forms of reception. Moreover, it questions the pre-existing classification systems for digital libraries and museums and argues for a more participatory model. Finally, it investigates the possibilities of creating information commonses based around diverse voices that indigenize processes of tagging, blogging, and commenting.

FULL PAPER SUBMISSION 32:

Reframing Book Publishing in the Age of Networking

AUTHOR(S):

Bradley, Jana
Vokac, Heather

ORGANIZATION(S):

University of Arizona, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

This paper presents preliminary results of an on-going examination of book publishing practices emerging through the complex interaction of technological, economic, and socio-cultural factors in the networking environment of the Internet. The theoretical framework guiding the study is a diachronic definition of book publishing proposed by Thomas R. Adams and Nicolas Barker in “A New Model for the Study of the Book,” first published in *A Potency of Life: Books in Society*, by the British Library in 1993. The Adams/Barker definition of publishing focuses on “the initial decision to multiply a text or image for distribution.” In this paper, we propose what we intend as a friendly amendment to their definition: for our purposes, a book publisher is an individual or a collective that makes the initial decisions and arrangements for multiple copies of books to be publicly available for distribution. The methodology for this work was to study a purposive sample of book publishers found on the Internet that fit our definitional framework. Our final sample, which we call emerging publishers, is just under 300 publishers. This sample was divided into three categories: Category I: Book Publishing Entities; Category II Author as Publisher and Category III Channels to Market. Each category is sub-divided, defined and described. Tables are included which show the publishers in each category. The paper concludes with observations across categories about format, shifts in publishers’ roles, standards of publishing practice, costs, and discovery, reception and reading and survival.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

Reframing Book Publishing in the Age of Networking

Jana Bradley, PhD
University of Arizona
1515 E First St
Tucson, AZ 85717
+1 (520) 621-3565

janabrad@email.arizona.edu

Heather Vokac, MLS
University of Arizona
1515 E First St
Tucson, AZ 85717
+1 (520) 621-3565

vokach@email.arizona.edu

ABSTRACT

This paper presents preliminary results of an on-going examination of book publishing practices emerging through the complex interaction of technological, economic, and socio-cultural factors in the networking environment of the Internet. The theoretical framework guiding the study is a diachronic definition of book publishing proposed by Thomas R. Adams and Nicolas Barker in “A New Model for the Study of the Book,” first published in *A Potency of Life: Books in Society*, by the British Library in 1993. The

Adams/Barker definition of publishing focuses on “the initial decision to multiply a text or image for distribution.” In this paper, we propose what we intend as a friendly amendment to their definition: for our purposes, a book publisher is an individual or a collective that makes the initial decisions and arrangements for multiple copies of books to be publicly available for distribution. The methodology for this work was to study a purposive sample of book publishers found on the Internet that fit our definitional framework. Our final sample, which we call emerging publishers, is just under 300

publishers. This sample was divided into three categories: Category I: Book Publishing Entities; Category II Author as Publisher and Category III Channels to Market. Each category is sub-divided, defined and described. Tables are included which show the publishers in each category. The paper concludes with observations across categories about format, shifts in publishers' roles, standards of publishing practice, costs, and discovery, reception and reading and survival.

Topics

Community technologies and networking
Cultural information systems
Information technology and services

Keywords

Electronic publishing
Publishing in the 21st Century
Emerging Publishers

71. INTRODUCTION

This paper presents preliminary results of an on-going examination of book publishing practices emerging through the complex interaction of technological, economic, and socio-cultural factors in the networking environment of the Internet. The theoretical framework guiding the study is a diachronic definition of book publishing proposed by Thomas R. Adams and Nicolas Barker in "A New Model for the Study of the Book," first published in "A Potency of Life: Books in Society, by the British Library in 1993 and reissued in 2001 in paperback by the British Library and Oak Knoll Press. They were interested in developing a definition that would accommodate changes in publishing practices from the manuscript book period to the present day. They comment that while the most familiar form of publishing is commercial, models of publishing that are not driven by profit have existed from the rise of the western manuscript book trade forward. They give as examples the publishing activities of governments, religious institutions, and individuals supported by patronage.

The Adams/Barker definition of publishing focuses on "the initial decision to multiply a text or image for distribution," (Adams/Barker, 2001, p. 15), and identifies four roles involved in this decision: author, financier, manufacturer and distributor, which in any given period of time might be performed by distinct parties or might be done in any combination. The Adams/Barker model also depicts five events in the life of a book: publication, manufacture,

Copyright and Disclaimer Information

The copyright of this document remains with the authors and/or their institutions. By submitting their papers to the iSchools Conference 2008 web site, the authors hereby grant a non-exclusive license for the iSchools to post and disseminate their papers on its web site and any other electronic media. Contact the authors directly for any use outside of downloading and referencing this paper. Neither the iSchools nor any of its associated universities endorse this work. The authors are solely responsible for their

distribution, reception and survival. These events are conceptualized as occurring within the context of what they call the socio-economic conjecture, including commercial pressures, intellectual, legal and religious influences, and the realms of social behavior and taste.

A diachronic definition of publishing, i.e. a definition that is generic enough to encompass the study of book publishing as it changes over time, is ideally suited to an examination of the changes that are occurring in the patterns and practices with which books today come into being as objects, are made publicly available, are discovered by potential audiences, read or not read, and survive or do not survive. In this paper, we propose what we intend as a friendly amendment to Adams/Barker: for our purposes, a book publisher is an individual or a collective that makes the initial decisions and arrangements for multiple copies of books to be publicly available for distribution.

It is worth noting that this definition, while we believe it gets at the essential activity underlying the diachronic practice of book publishing, does not require features of the publishing practices we know best from the twentieth century, such as commercial mainstream publishing companies with large financial investment for manufacture, tight gate keeping of published material, a body of book writing and publishing expertise interpreted and enforced by editors and publishers, consolidated distribution, and above all, print format. All these features certainly still exist in publishing, and arguably, still dominate, but by using a definition that does not require these features, we can not only see continuity with publishing from the manuscript book forward, but we are more likely to see changes that are emerging in the technological, economic and socio-cultural world of networking as currently instantiated on the Internet. In addition to their diachronic

definition of book publishing, the theoretical context of Adams and Barker's work, including the identification of historically shifting roles and the socio-economic conjecture in which they situate the

activities and technologies that comprise the events in the lives of books, shapes our thinking in this study.

The focus of this study is book publishing, although the diachronic definition of publishing above can be applied to other types of content. For our purposes, we have defined a book loosely as any publication of monographic length with a sustained narrative or argument, either textual or visual or in combination, and some familial relationship to the genre we

Category I: Book Publishing Entities: Entities (or individuals functioning as entities) who arrange to make books from multiple authors available electronically
I.1 Publishers of born-print books providing electronic distribution (print to digital)
I.2 Publishers of born-electronic books, some providing POD print distribution (digital to print)

Category II: Author as Publisher: authors themselves making the initial decision to publish their own work
II.1 Using publishing services
II.2 Using automated publishing websites
II.3 Using author-created and controlled website infrastructure

Category III: Channels to Market: Third-parties assisting publishers of both types with a variety of publishing functions.

know as the printed book, including some combination author, chapters, tables of contents, ISBN number and other resemblances.

72. METHODOLOGY

The methodology for this work was to study a purposive sample of book publishers that fit our definitional framework, continuing to search until we were not retrieving examples that seemed new in significant ways. Our final sample, which we call emerging publishers, is just under 300 publishers. Given the enormity and constantly changing nature of the Internet, this is obviously not a complete universe. We also chose not to examine print publishers who did not use themselves use the Internet for distribution and sales of books, although many of these publishers' works are available through the distributors and aggregators that were included. The ways that digital technologies are changing traditional print publishing practices, excluding electronic books, have been documented thoroughly by John Thompson in *Books in the Digital Age* (2005). Those traditional publishers who are experimenting with the emerging publishing environment on the Internet are represented well in our sample.

73. CATEGORICAL ANALYSIS

Our goal was to organize our sample around categories whose characteristics call attention to key differences among our publishers and their activities. The categories were not intended to be mutually exclusive but to highlight both shared characteristics and some different patterns within categories. After much experimentation, we identified three types of publishing agent, two who make the initial decision and arrangements to publish, and a third category used by both types of primary publishers for publishing arrangements. The first category, book publishing entities, was defined as entities (or individuals functioning as entities) who make initial decisions to make books from multiple authors publicly available. The second category, authors as publishers, was defined as authors themselves making the initial decision to publish their own work. The third category was defined as groups that assist publishers of both types with a combination of publishing roles. This group was labeled "channels to market," following increasingly common usage among sites studied. Categories were then subdivided, again along common characteristics. The intention is not to make a rigid classification scheme, but to highlight distinctions in a systematic way.

Table I: Categories of Emerging Publishers

Our descriptions of these categories in the text below follow roughly the same pattern: a label and definition that attempts to capture defining feature of each category; a description of common characteristics; and then a brief discussion of a handful of examples that illustrate either characteristic or distinctive features within a category. Tables accompanying each discussion show the publishers in each category. At the end of the descriptions of categories, we bring together some

observations about shifts in book publication practices resulting from the analysis of the sample across categories.

73.1 Category I: book publishing entities:

This category comprises entities who make the initial decisions for making multiple copies of books from multiple authors publicly available. Most of these entities appear to be commercial, for-profit companies.

It is useful to think of these publishers in two groups: (I.1) those whose primary production of books is print and who then may make decisions to distribute electronically; and (I-2) those whose primary production of books is electronic and who then may make decisions to distribute in print POD format;

Publishers of Born-Print E-Books: The key characteristic highlighted by this category is the electronic offering of books that first appeared in print. Our sample includes fifteen print publishers with substantial electronic distribution announced on their sites, by no means all that exist but enough of a variety to show some common characteristics and some different arrangements. As with born-print e-journals, born print e-books bring with them all the standard value-added services of a print publisher; they are simply distributed electronically. Format of distribution and availability, along with related changes in discovery mechanisms, are features distinguishing this category of electronic publisher. In other respects, they follow commercial print publishers publishing practices, including the acquisition and preparation of manuscripts and the compensation of authors. Having said that, it is highly likely that the pressures and competition of the emerging publishers discussed below will over time change more print publishing practices, as it is clearly influencing distribution and marketing now.

Simon and Schuster E-Books serves as an example for this category:

<http://www.simonsays.com/content/index.cfm?pid=523087&tab=6>. It is a section of the Simon and Shuster website; it shows covers, blurbs, and publication information, and available formats. At present, Simon and Schuster e-books are downloadable in Adobe, Microsoft and Palm e-book formats. eHarlequin.com eBook Boutique, <http://ebooks.charlequin.com>, and ePenquin, <http://www.penguin.co.uk/>, follow a similar pattern.

Several other models can be seen within this group. Norton eBooks.com, <http://www.nortonebooks.com/welcome.asp>, offers two options for electronic books: reading for a year online, purchased through Norton, or downloading through [Powells.com](http://www.powells.com), a vendor of e-books in Adobe, Microsoft and Palm as well as new and used traditional print books. Cambridge University Press has a site, Cambridge eBookstore, which serves as a catalog of Cambridge University Press eBook offerings. Purchase and download of e-books is handled through a distributor, eBooks.com. Hachette Book Group USA, for book publishers owned by Hachette, <http://www.hachettebookgroupusa.com/>, serves as a "union catalog" for its publishers. A "Where to Buy" link

has extensive lists of booksellers' websites, audio book vendors and online retailers.

Table II provides a list of publishers of born-print e-books from our sample.

Table II: Publishers of Born-Print E-Books

- Cambridge University Press eBookstore
<http://www.cambridge.org/ebookstore>
- Del Rey online <http://www.randomhouse.com/delrey/ebooks>
- eHarlequin.com eBook Botique <http://ebooks.charlequin.com>
- ePenguin <http://tinyurl.com/2h92o3>
- Franklin.com <http://www.franklin.com/about/profile.asp>
- Hachette Book Group, USA
<http://www.hachettebookgroupusa.com>
- The Hindu eBooks <http://www.thehinduebooks.com>
- Imagine eBooks <http://imagine-ebooks.com>
- MTB Europe <http://www.mtbeurope.info/books/ebooks.htm>
- Mythic Island Press <http://www.mythicisland.com>
- NortonEbooks.com <http://www.nortonebooks.com/welcome.asp>
- Simon & Schuster
<http://www.simonsays.com/content/index.cfm?pid=523087&tab=6>
- Tanner Ritchie Publishing <http://www.tannerritchie.com>
- Taylor & Francis eBookstore
<http://www.ebookstore.tandf.co.uk/html/index.asp>
- White Wolf <http://www.white-wolf.com>

Publishers of Born-Digital E-Books: This category features publishers of electronic books that have not been preceded by print books, in other words, born-digital e-books. We found over 90 examples of publishers that fit this definition. In general, many patterns and practices of these publishers resemble those of print publishers. These publishers make editorial decisions in determining which manuscripts to select, either from open FULL PAPER SUBMISSION or by invitation; they do not charge authors any fees for publishing; they have contracts specifying royalties and rights; they provide the typical value-added services of print publishers including arranging for the publishing infrastructure, financial underwriting, quality control editorial and design services, and marketing through their website, acting like an online bookstore or catalog.

Publishers in this category distributed multiple copies of original work made publicly available for the first time primarily through a variety of electronic means, in a variety of formats for viewing on the consumer's computer or on a variety of viewing devices, with a variety of controls over use. A brief discussion of e-book formats, reading devices and digital rights management (DRM) is necessary here, but will inevitably be out of date quickly, since the volatile and rapidly changing market is conditioned by multiple business arrangements among publishers, channels to market, and device manufacturers. An industry standard has yet to emerge, and some argue that the difficulty of choosing formats and actually downloading to devices, rather than an aversion to reading on screens or a predilection for the print book as an object, has prevented an explosion of consumer

preference for electronic over print. In December 2005, Michael Hyatt, CEO and publisher of Thomas Nelson Publishers, in a blog posting titled "The Death of the Traditional Book," popularized the now oft-repeated phrase, "one device away" from a digital book revolution. Some have suggested that Amazon's Kindle, discussed below, is, if not the one device that will revolutionize digital books, at least a harbinger of that device.

The overview of e-book formats and devices from Fictionwise (discussed below under the category of channels to market, <http://www.fictionwise.com/help/help.htm>), is reproduced below as an indicator of the variety of formats and devices now available:

- Sony Reader (Sony Reader PRS-500 and PRS-505 [.lrf])
- Kindle (Kindle [.mobi])
- Personal Computers (Microsoft Reader [.lit], Adobe Acrobat [.pdf], Mobipocket [.prc])
- Palm OS Handhelds (Palm Doc [.pdb], iSilo [-is.pdb], Mobipocket)
- Pocket PC Handhelds (Microsoft Reader [.lit], Mobipocket [.prc])
- Win CE Handhelds (Palm Doc [.pdb], Mobipocket [.prc])
- Rocket/Reb1100 Devices (rocket files only [.rb])
- Franklin eBookMan (Mobipocket, eBookMan files [.fub])
- Hiebook (Hiebook [.kml])
- Symbian OS Organizers and Smartphones (Mobipocket [.prc])
- Other Devices (many can read Palm Doc)

Fictionwise describes their electronic offerings as either multi-format or secure format. The key difference here is encryption or digital rights management. The Fictionwise discussion of this difference

((<http://www.fictionwise.com/help/ebook-formats-FAQ.htm#difference>)) is an excellent introduction to devices, formats and digital rights management, and their FAQ offers some insight into the business decisions that shape the conditions of availability.

Many of these publishers also offered POD print versions, either at their discretion depending on sales, or less frequently, at the author's choice. It is common for these publishers to use third parties for both manufacture and simultaneous distribution of print copies. Most books have ISBN numbers. In general, most of these publishers do not pay advances but royalty percentages could be larger than print percentages, keeping in mind that costs per copy was often less than print.

Although many publishing practices of these emerging publishers are similar to patterns from commercial print

publishing, it is important to recognize ways in which these emerging publishers are pushing the traditional boundaries of publishing established in the print world. Content is one of the most obvious differences. Fiction, and particularly romance and other genre fiction, dominated our sample, although non-fiction was available. Imprints, or series with specific characteristics, are frequently used to direct niche audiences to their work. The categorical lists which comprise the primary subject finding tool for all categories of publishers show some categories that arguably have not been well represented across print publishing, such as erotica, dark fantasy, and others.

Lower entry costs for these publishers, compared to traditional print publishing, may account for the flourishing of niche publishing, publishers and imprints within publishers that are targeted to specific, and frequently non-mainstream, audiences.

Double Dragon Publishing, <http://www.double-dragon-ebooks.com/index.php>, is an example chosen for this category. Founded in 2001, it is primarily an e-book publisher, although it chooses some titles for POD print distribution through lulu.com, especially in large print versions. It provides 30% royalties paid on a 6-month cycle; it buys electronic rights and copyright is retained by author or institution. It does not charge fees to authors. According to their website, 90% of their sales come from Science Fiction, Fantasy, Horror and Speculative Fiction.” In 2004, they opened a romance line. Services provided to authors include editing services, cover and banner design, free copies for use in promotion, and an authors-only forum. In addition to a long list of genre fiction, they have more than 10 titles in the non-fiction category and also in the self-help category. They provide e-books in the following formats: Rocket-eBook, Hiebook, Adobe PDF, MS-Reader, Mobipocket, iSilo, Franklin eBookMan, and Palm Doc, and the website assures readers that they will keep up with new technologies.

New Concepts Publishing, <http://www.newconceptspublishing.com/authorfaqs.htm>, another example, was the first totally electronic romance publishing house. Like most other examples in this category, NCP assigns ISBN numbers, pays royalties (30%-40%), provides cover art, and buys exclusive electronic rights for a specified amount of time. NCP decides, based on interest and sales, whether a title will be printed as a POD or trade paperback. E-books are provided in RTF, HTML, MS Reader, MobiPocket, and PDF.

Writer's Exchange E-Publishing differs from the two examples above in that authors have the option of paying set-up print fees and having the book appear in print as well as electronically. Writer's Exchange publishes e-books which they distribute through their online store, Reader's Eden. They also distribute electronic books by other authors.

Awe-Struck E-Books, <http://www.awe-struck.net/>, publishes genre fiction and provides an example of niche publishing with their Ennoble imprint, claiming to be unique in featuring heroes and heroines with disabilities.

Table III provides an incomplete list of publishers of born-digital e-books in our sample. Please email authors for complete list.

Table III: Publishers of Born-Digital E-Books

- Alt-X Ebooks http://www.altx.com/ebooks/altx_frame.html
- Amber Quill Press (Also Amber Heat and Amber Allure imprints) <http://www.amberquill.com>
- Amira Press <http://www.amirapress.com>
- Aphrodite's Apples <http://www.aphroditesapples.com>
- Arch eBooks <http://www.archebooks.com>
- Artemis Press <http://www.artemispress.com/html/e-bookstore.html>
- Asylett Press <http://www.asylett.com>
- Atlantic Bridge *includes nonfiction <http://www.atlanticbridge.net>
- Awe-struck ebooks <http://www.awe-struck.net>
- Baen Books <http://www.baen.com>
- BellaOnline <http://www.bellaonline.com/ebooks>
- Belgrave House <http://www.belgravehouse.com>
- Best Buy Ebooks (only two authors, Print through lulu) <http://store.buy-best-ebooks.com>
- Books for a Buck <http://www.booksforabuck.com>
- Books Unbound <http://www.booksunbound.com>
- BookShaker *nonfiction only <http://www.bookshaker.com/index.php>
- Bosun Books <http://www.cmonline.com/bosun>
- Bowbridge Publishing <http://www.bowbridgepublishing.com>
- Carnal Desires Publishing <http://www.carnaldesirespublishing.com>
- Champagne Books <http://www.champagnebooks.com>
- Changeling Press Erotic Fiction <http://www.changelingpress.com>
- Chippewa Publishing LLC <http://www.chippewapublishing.com>
- Cobblestone Press <http://www.cobblestone-press.com>
- Computer Classics <http://www.computer-classics.com>
- Connexions *nonfiction <http://cnx.org/lenses/cnxorg/books>
- Creative Guy Publishing <http://www.creativeguypublishing.com>
- Cybereditions *nonfiction <http://cybereditions.com>
- Dark Eden Press <http://www.darkedenpress.com>
- Derlinger's Publishers *includes nonfiction <http://www.thebookden.com>
- DiscoverABook.com <http://www.discoverabook.com>
- DiskUs Publishing <http://www.diskuspublishing.com>
- DLSIJ Press *includes nonfiction <http://dlsijpress.com>
- Double Dragon Publishing <http://www.double-dragon-ebooks.com>
- Dpdotcom.com <http://www.dpdotcom.com>
- Dragonfly Publishing <http://www.dragonflypubs.com>
- Drollerie Press <http://www.drolleriepress.com>
- Eastgate <http://www.eastgate.com>

- Ebooksonthe.net <http://www.ebooksonthe.net>
- Ebooks on Disk
<http://www.ebooksondisk.com/index.php>
- Ebooksonthe.net *also sells previously published books
<http://www.ebooksonthe.net>
Incomplete list. Please email authors for complete list.

73.2 Category II: author as publisher:

The second category highlights the author acting as publisher: the person who makes the initial decisions and arrangements to make multiple copies available for public distribution. Digital technologies have opened many new avenues for authors to create books and make them publicly available, and authors are using them in great numbers. To provide an overview of the burgeoning world of authors as publishers, we have organized these efforts in three sub-categories. These categories, even more than the ones introduced previously, are an oversimplification of the great variety that exists in these efforts, but they have the virtue of bringing some order to an examination of this phenomenon. The three categories examined here are authors using fee-based commercial services, authors using automated websites designed for uploading books, and authors providing their own web infrastructure for making their work available, and sometimes sharing their space with other authors.

73.2.1 Authors using fee-based publishing services

Adams and Barker point out that authors arranging for the publishing of their own work, often with the assistance of patronage, has a long history. In the twentieth century, author-financed publishing carried the pejorative title of vanity publishing, with the implied assumption that if authors were any good, they would not have to finance the publication of their own works. Authors who arranged for the manufacture of their own printed books faced major obstacles in distributing them. In the pre-Internet print book world, distribution of books was in the hands of a few wholesalers, who supplied retailers, libraries and schools. Amazon pioneered the “channel to market” approach on the Internet, providing for the discovery, distribution and sales of print titles, soon expanded to media and other commodities, and thus breaking up the distribution bottleneck.

In early January, 2008, the Associated Press distributed an article by Candace Choi, “Got a Manuscript? Publishing Now a Snap,” highlighting the skyrocketing use by authors of fee-based commercial publishing services. Our searching identified over fifty of these services, of which Lulu.com, Amazon’s CreateSpace, and Blurb.com were featured by Choi. The basic idea behind these services is that authors can purchase publishing infrastructure, manufacturing, distribution and marketing services previously provided for free by print publishers for their selected authors, shifting both the decision to publish and the financial burden to the author, but giving them opportunities to reach audiences previously dominated by commercial companies. While the main thrust of most services is to provide print books, many

provide electronic distribution as well. Two examples of fee-based publishing services will be discussed: iUniverse and Amazon’s CreateSpace.

iUniverse is a useful example to describe briefly because they have organized their services into packages, ranging from under \$400 to under \$1400, with add-ons in all categories. The most inexpensive option is most like traditional vanity publishing of the past, providing the author with professionally manufactured POD print copies for family and friends, with the added limited distribution mechanism of availability through the iUniverse book store. Increasingly expensive options provide more professional quality and services, and increasingly wide retail availability through channels to market, such as Barnes and Noble.com, Amazon and many others. The high-end services provide the essentials of commercial print publishing, including marketing efforts such as the distribution of review copies.

Amazon’s CreateSpace calls itself “on demand” publishing. They manufacture and sell physical objects, including POD books, audio books, audio CDs, video CDs and other products one at a time, as ordered. The price is set by the author and royalties are paid to the author as they are generated by sales. The author or creator is entirely responsible for the content which is uploaded to CreateSpace for production. CreateSpace products are sold via Amazon. While theoretically the consumer can identify a self-published book by recognizing the publisher, the increasing number of these services, plus the increasing mix on channel to market sites of commercial publishers, traditionally printed books, and emerging publishers of POD books and electronic books of many varieties means that in practice, the distinction between commercial publication and author-subsidized publication is increasingly blurred.

Table IV provides a list of the fee-based publishing services in our sample.

Table IV: Fee-Based Publishing Services

- Antony Rowe Ltd. <http://www.antonyrowe.co.uk>
- AuthorHouse <http://www.authorhouse.com>
- Authors Online <http://www.authorsonline.co.uk>
- Aventine Press <http://www.aventinepress.com>
- Blitz Print
<http://www.blitzprint.com/index.htm?aim=c01k002g>
- Booklocker <http://www.booklocker.com>
- Bookstand Publishing <http://www.ebookstand.com>
- Booksurge <http://www.booksurge.com>
- CafePress
<http://www.cafepress.com/cp/info/sell/products/books>
- Cheap eBooks *eBook POD
<http://www.cheapebookshop.com>
- CreateSpace
- Dog Ear Publishing
<http://www.dogearpublishing.net/getstarted.aspx>
- Ebookomatic (operates more like a POD but produces ebooks only) <http://www.ebookomatic.com>
- EbooksLib *also multilingual, also distributor

<http://www.ebookslib.com>

- Equilibrium Books <http://www.equilibriumbooks.com>
- First Choice Books <http://www.firstchoicebooks.ca>
- Foremost Press <http://www.foremostpress.com/authors>
- Fultus Publishing <http://www.fultus.com>
- GASLight Publishing (traditional and subsidy pub)
<http://gaslightpublishing.net>
- GLB Literary Corner *includes nonfiction
<http://www.glbpubs.com>
- Golden Pillar Publishing
<http://www.goldenpillarpublishing.com>
- Illumina Press <http://www.illumina.com>
- Infinity Publishing <http://www.infinitypublishing.com>
- Inkwater Press <http://www.inkwaterpress.com>
- Instabook Publisher
<http://www.instabookpublisher.com/files/faq.htm>
- iUniverse <http://www.iuniverse.com>
- Ka-Blam <http://www.ka-blam.com>
- Keystone Digital Press
<http://www.greenepublicationsinc.com>
- Lightning Source <https://www.lightningsource.com>
- Lulu.com <http://www.lulu.com>
- Net Publications Inc.
<http://www.netpub.net/newservices/pubserv.asp>
- Outskirts Press <http://www.outskirtspress.com>
- Page Free Publishing <http://www.pagefreepublishing.com>
- Pleasant World – Christian Print on Demand Self
Publishing <http://www.pleasantword.com/Default.asp>
- Poseidon Books <http://www.poseidonbooks.com>
- Private Ice Publications <http://www.skatefic.com>
- Publish America <http://publishamerica.com>
- Publishers Row *fee based epub
<http://www.publishersrow.com>
- Star Publish, LLC <http://starpublish.com>
- Tate Publishing <http://www.tatepublishing.com>
- Trafford Publishing <http://www.trafford.com>
- U-Publish <http://www.u-publish.com/testbook.htm>
- Universal Publishers <http://www.universal-publishers.com>
- Unlimited Publishing <http://www.unlimitedpublishing.com>
- Virtual Bookworm <http://www.virtualbookworm.com>
- Wheatmark <http://www.wheatmark.com>
- Wing Span Press <http://www.wingspanpress.com>
- Writers Print Shop
<http://www.writersservices.com/wps/index.htm>
- Writersworld <http://writersworld.co.uk>
- Xargol *Hebrew and English
http://www.xargol.com/index.php?cat=2&name=o_fiction
- xLibris <http://www2.xlibris.com/pubservices/index.asp>
- Xulon Press <http://www.xulonpress.com>

73.2.2 Authors using automated websites to upload books
This category illustrates automated sites that exist solely to provide mechanics for authors to upload and share their work, free for the author and free for the reader. Our sample

includes sixteen of these sites. Mechanics, conditions, registration, and membership all differ. Although sites vary, many have a high proportion of uncompleted or abandoned work. Four examples will be highlighted: FictionCentral.net, NaNoWriMo, Writing.com and Buried.com.

FictionCentral.net, <http://www.fictioncentral.net/index.php>, describes itself as a “place to write and read original fiction stories.” Authors need to register; readers are encouraged to register and are requested to leave a review of what they have read. Works can be posted chapter by chapter, and a notation indicates when it is completed. The site administrators have articulated rules for writing and reviewing, including a system of ratings and warnings, including the requirement to rate material as 12+ for people over 12, and the notation of SLASH for material including single-sex pairings, and courtesy rules such as no rudeness or swearing.

NaNoWriMo, <http://www.nanowrimo.org/>, is an engagingly original attempt to encourage people to jump in and write a novel. In November, registrants have one month to complete a 50,000 word novel. The NaNoWriMo website says “It’s all about quantity not quality.” Unlike many other sites to upload work, the author’s work on NaNoWriMo work remains private unless the author posts excerpts or shares with others. At the end of the month, an author’s word count can be verified and if it exceeds 50,000, the author is listed as a winner.

Writing.com, <http://www.writing.com/>, describes itself as “the online community for writers and readers of all interests and skill levels.” It requires registration, in which the registrant provides a username that will be used as his/her online identity, and required personal information for the site administrators. Many features include space for authors’ writing profiles and many interactive community venues. Buried.com is an example of a site containing content around a theme, in this case, “everything that is horror...” and includes mechanisms for authors to upload their own writing, again in this case horror fiction.

Table V lists the automated publishing websites in our sample.

Table V: Automated Publishing Websites

- ABC Tales <http://www.abctales.com>
- Afterglare <http://afterglare.com/browse.php?type=titles>
- Badosa <http://www.badosa.com>
- eStories *German, English, Spanish, Dutch, French, Italian, Portuguese <http://www.e-stories.de>
- Fanstory.com <http://www.fanstory.com>
- FictionCentral <http://www.fictioncentral.net>
- Fiction Press <http://www.fictionpress.com>
- FicWad <http://www.ficwad.com>
- NaNoWriMo <http://www.nanowrimo.org/>
- NoviceWriters <http://www.novicewriters.net/index.html>
- Shur’tugal
<http://fanfiction.shurtugal.com/categories.php?catid=12>
- Slash Cotillion <http://cotillion.slashcity.org>
- UseNet (Google Groups) <http://groups.google.com>
- Writing.com <http://www.writing.com>

- Buried.com <http://www.buried.com/features/novels.shtml>
- Cosmos Magazine Fiction Serials
<http://www.cosmosmagazine.com/fiction/online/serials>

73.2.3 Authors creating their own infrastructure to publish their work

The relative ease of website construction has enabled numerous authors to create sites as venues for their own work. Some sites exist solely for the author's content, with some or all items free to consumers, and some sites exist primarily for the author's work but accept works from other writers, again with variation in fee or free. Since a great deal of variety exists here, those interested should explore the lists of sites created and controlled by authors in Table VI.

Table VI: Websites Created and Controlled by Authors

Single-author only

- Anisky's Abode
<http://www.geocities.com/melsstories/index.html>
- Back in School: A Love Story
<http://www.alovestory.net/92.html>
- Backpacking on Little Money
<http://backpackingonlittlemoney.wordpress.com>
- Caring for God's Laptop
<http://www.important.ca/godslaptop>
- Check.org <http://www.check.org/fiction/index.htm>
- The Circle of Ceridwen
<http://octavia.net/books/booklist.html>
- Coyote
<http://www.hauntedhousedressing.com/coyote.htm>
- Dark Icon <http://www.darkicon.com/Library>
- Darkfin <http://www.darkfin.com/main.html>
- Diary of a Heretic <http://www.diaryofaheretic.blogs.com>
- D. S. Bauden Original Fiction
http://www.dsbauden.com/original_fiction.shtml
- Fiction by Junkfood Monkey
<http://www.junkfoodmonkey.pwp.blueyonder.co.uk/Fics/ficindex.html>
- Fields of Night <http://users.california.com/~sara Peyton>
- From in the Shadows
<http://fromintheshadows.blogspot.com>
- Furry Magic <http://www.furrymagic.org/disclaimer.htm>
- Hopcott ebooks <http://hopcott.net>
- If He Should Lose His Own Soul
<http://www.fables.co.uk/grownups.htm>
- Jackee's Corner <http://fiction.jackeescorner.com>
- Kuroiyusei Original Fiction
<http://www.kuroiyousei.net/Ofiction.html>
- Lady Jaída original fiction
<http://ladyjaída.livejournal.com/tag/original+fiction>
- Maximillian Lückenhaus *German http://www.wahre-kunst.de/Lueckenhaus/m_sfi.html
- Metropoppyfield
<http://www.metropoppyfield.com/buy.html>

- Moments of Existence
http://www.freewebs.com/moments_of_existence/home.html
- Mortal Ghost <http://mortalghost.blogspot.com>
- Nate Simpson
<http://www.natesimpson.com/writing/#Novels>
- Nicole West Original Fiction <http://www.nicolewest.net>
- Paddy – Original Fiction
http://www.originalpain.com/paddy/fiction_paddy.html
- Prose Dreams <http://www.prosedreams.net>
- The Relic Triangle <http://www.relictriangle.com>
- Richard A. Bartle Original Fiction
<http://www.mud.co.uk/richard/fiction.htm>
- Salamanders <http://home.earthlink.net/~salamanders/>
- ScribeScribbles
<http://www.angelfire.com/grll/scribescribbles/mainmenu.htm>
- Setisia <http://www.freeonlinereading.com>
- Starving Writer <http://www.starvingwriter.com>
- The Talespinner
<http://www.angelfire.com/home/talespinner>
- The Wonderful World of Bradley Stoke
<http://bradleystoke.0catch.com>
- Xenacast Original Fiction
<http://www.xenacast.com/xc/originalfiction.htm>
- Zvi's Palm Jewish Fiction Page
<http://home.att.net/~fiddlerzvi/PalmLit.html>

Single Authors and Sharing:

- B.L. Miller's Site <http://blmiller.net>
- Free Novels Online <http://starry.com/novel/authors.htm>
- Shoujo-Ai Canada
<http://www.katandnekomanga.ca/writing/shoujo-ai.htm>
- A Room of Her Own: A Dynamic Anthology of Lesbian Fiction <http://blmiller.net/room/index.html>
- Aphelion Webzine <http://www.aphelion-webzine.com>
- The Cobweb's Short Stories
<http://www.angelfire.com/ego/thecobwebs/stories.htm>
- Dark Lethe <http://reactivewriting.com/dark-lethe>
- Elfwood <http://www.elfwood.com/libr/entrance.html>
- Free the Writer
<http://www.freethewriter.com/Archives.html>
- Ideomancer
<http://www.ideomancer.com/main/ideoMain.htm>
- Iyinka's Original Fiction

73.3 Category III: channels to market: third parties who acquire rights to manufacture, distribute and sell print or electronic books

Channels to market, as mentioned in the definition of categories, is the term increasingly used to represent the

combination of manufacture, distribution, discovery and sales mechanisms that is flourishing on the Internet. In the pre-Internet print book world, distribution of print was tightly controlled by a few book wholesalers, who supplied titles from multiple publishers to sellers and institutions, such as schools and libraries. Discovery of titles was accomplished primarily through publishers advertising and marketing, reviews, compendia such as Books in Print, library catalogs and browsing bookstore shelves. Amazon pioneered the “channels to market” concept of providing a mechanism for consumers to find and order print titles from multiple publishers, extending to used as well as new books and into other media, including electronic books. “Channels to market sites,” contract with publishers and sometimes publishing services to sell print and electronic books on their sites in specific formats, readable on specific devices.

Initially, we had classified channels to market as a secondary type of commercial publishing entity, along with primary born-print and born-digital publishers of electronic books. However, toward the end of our analysis, we put them in their own category because of the increasing tendency of both commercial publishers and authors who use publishing services to be combined in channels to market. Not all channel to market sites include books from publishing services, but Amazon as one of the largest channels to market companies does, and for numerous other publishing services, it is a fee-based alternative authors can choose. As a result, channels to market are increasingly including titles from a mix of publishing sources. In other words, authors who publish their own work can be included in some of the most used channels to market.

Twenty-seven “channel to market” sites were identified in our sample. Several examples have been chosen for short descriptions: eBooks, Fictionwise, NetLibrary and Overdrive, and Mobipocket and Kindle Store (both Amazon companies), and LightningSource.

eBook.com calls itself “the digital bookstore.” It sells electronic copies of print titles and also born-digital titles. It contracts with publishers to buy a non-exclusive Internet right to digitally store, reproduce and transmit titles; royalties are paid to the author through the publisher. e-Books distributes books in Adobe PDF and encrypts books using Adobe Content Server.

Fictionwise, an e-book and audio-book publisher <http://www.fictionwise.com/AuthorInfo.htm> mentioned earlier, is interesting because in addition to contracting with publishers, it will contract directly with authors who have more than 10 published print books. It offers books in unencrypted formats and in Adobe, Microsoft and Mobipocket Secure (i.e. encrypted) formats.

NetLibrary and Overdrive are examples of companies that do not sell primarily to consumers but to third parties.

NetLibrary is the eContent division of OCLC. When accessing NetLibrary from a member library system, the NetLibrary site shows the titles owned by the member. Overdrive targets multiple third parties, including libraries, schools, retailers, and enterprises and tailors the product it

sells to each market. Overdrive also hosts the Digital Media Locator, <http://www.ebooklocator.com/>, which locates libraries that have contracted with Overdrive to provide patrons with free downloads of specific titles. The Locator is a strategic alliance with OCLC, utilizing Worldcat. Mobi-Pocket (since 2005 an Amazon company) and Amazon are examples of channel to market sites that have their own branded devices but also act as channels for content for other devices. Amazon has recently introduced the Kindle, which has many of the features Hyatt mentioned as necessary for the “one device” that would spark consumer interest in digital books. Kindle downloads digital content, including books, newspapers, magazines, and blogs, from the Kindle Store, <http://www.amazon.com/b?ie=UTF8&node=133141011>, either directly from your Kindle or from the Amazon site to your Kindle via wireless. The device itself is the size of a paperback book. Experts and consumers alike have varying opinions of reading quality. The Fictionwise Kindle FAQ offers insights into Amazon’s business decisions with regard to channels other than Amazon for distributing to the Kindle and illustrates the jockeying for position that now characterizes the e-book market. At this writing in late January, 2008, the Amazon site lists the Kindle as temporarily out of stock, but is taking orders.

LightningSource, <https://www.lightningsource.com/>, describes itself as “demand-driven book manufacturing and distribution” services. They provide rapid print-on-demand electronic book fulfillment services for publishers to retailers and libraries. LightningSource is illustrative of the emergence of a stream of print POD books from either print or electronic files supplied by the publisher, as well as electronic books from digital files provided by the publisher. Interestingly but perhaps not surprisingly, LightningSource is a subsidiary of Ingram Industries, owners of Ingram Book Group, one of the largest wholesalers of books in the United States. LightningSource does not reveal their criteria for what and who constitutes a publisher, but it is highly probable that small niche publishers, even individuals acting as publishers, can enter the publishing arena through LightningSource. According to the LightningSource website, they print a million books a month and the average print run is 1.8 books.

Table VII: Channel to Market Sites

- Fictionwise
- ACLS Humanities E-Book <http://www.humanitiesebook.org/intro.html>
- Adobe Digital Media Store <http://www.adobe.com/products/digitaleditions/library>
- Amazon.com (also distributes born digital ebooks) <http://tinyurl.com/33xrm2>
- Aneas eBookStore <http://www.aneas.net/ebookstore>
- Audible.com *audio books <http://www.audible.com>
- Books on Board <http://www.booksonboard.com>
- Bookshare.org *ebooks for the visually impaired <http://www.bookshare.org>
- Campus eBookstore <http://www.campusebookstore.com>
- Cartoon eBooks <http://www.cartoonebooks.com>

- Cyberread <http://www.cyberread.com>
- Diesel Ebooks <http://www.diesel-ebooks.com>
- eBook Mall <http://www.ebookmall.com>
- ebooks.com <http://www.ebooks.com>
- eBooks from Tess
http://www.tessebooks.com/index.php?cPath=25_10
- EbooksLib *also multilingual, also POD
<http://www.ebookslib.com>
- eFollett.com <http://ebooks.efollett.com>
- eLibrary <http://e-library.net>
- eReadable.com <http://www.ereadable.com>
- eReader.com <http://www.ereader.com>
- MemoWare PDA Bookstore
<http://www.pdabookstore.com>
- Mobipocket <http://www.mobipocket.com>
- NetLibrary <http://company.netlibrary.com/AboutUs.aspx>
- Numilog eBooks <http://www.numilogebooks.com>
- Powell's Books
<http://www.powells.com/ebookstore/ebooks.html>
- TecKnoQuest <http://www.tecknoquest.com>
- OverDrive <http://www.overdrive.com/>

74. OBSERVATIONS ABOUT SHIFTS IN BOOK PUBLISHING PRACTICES IN THE AGE OF NETWORKING

A number of observations can be made about emerging shifts in publishing practices by looking across the categories analyzed above. Many of these observations are not new to this paper, but they have arisen here from empirical examination of a large number of instances, and therefore they provide some evidence toward validation of these shifts.

74.1 Formats

While it is startlingly clear that the appearance of books in electronic format, either born-print or born-digital, and indeed the burgeoning number of new born-digital publishers, is a major phenomenon occurring on the Internet, focusing only on electronic books masks the reality that re-formatting goes in numerous directions, certainly from print to various electronic forms, but also to audio books, and from digital to print, enabled by POD.

The growing quality and popularity of POD, and its growing transparent use by publishers, either directly or through services like LightningSource, and channels to market like Amazon, is increasing the number of print titles available and diluting at least the traditional notion of the technological origins of a printed book and arguably the traditional print-publisher based notion of the book itself. The popularity of POD print copies for authors publishing their work through publishing services also illustrates that the flow of formats is complex. Some channels to market offer format as a consumer choice, and one may speculate that that trend may continue. Commercial channels now exist for the direct ordering of POD books by libraries, schools and retailers, although the extent of their use is not clear. The impact of

POD on book publishing is, we believe, in its infancy and will have possibly as dramatic a change on publishing patterns and practices as electronic books are having now.

74.2 Shifts in publishing roles

Our theoretical model arose out of notion that component roles of publishing have shifted over historical time from the rise of the western book and has directed our attention to dimensions of these shifts now occurring in the technological, economic and socio-cultural environment of the Internet. A central result of using the amended Adams/Barker definition of a publisher was a clearer understanding of the extent to which authors can act, and are acting, as their own publishers, and making available a wide variety of books, from products that can not be distinguished from the products of traditional publishers in appearance, and arguably in quality, to incomplete, informal, and unfinished explorations in writing books.

Another central recognition for us was the extent to which traditional publishers' patterns and roles in manufacturer, distribution and sales are changing. The rise of channel to market services which offer remanufacture, reformatting, multiple-format distribution directly to customers or through third parties, and new marketing approaches through interaction with authors and readers, while not eliminating publisher's decision-making certainly tends toward uncoupling or expanding functions that were once directly controlled by publishers.

74.3 Standards of publishing practice

Andrew Keen (2007), author of *Cult of the Amateur: How the Internet Is Killing Our Culture*, is just one of many voices who are concerned about the loosening of standards in publication and public discourse. Increasingly, the fact of publication is not a promise to the reader that the book has been vetted by the writing and editing professional elite who have dominated print publication, although many publishing services either offer or recommend the services of fee-based publishing professionals. The emergence of niche publishing, especially in areas that have not been well represented in mainstream publishing, either because of small market size or some form of stigmatization in mainstream society, can also be seen as a loosening of the gate keeping role of mainstream publishing in determining the body of book content available to the public. Not everyone decries this increasing inclusion of previously excluded content, both in substance and style, but the effects on society are certainly being vigorously debated.

74.4 Costs

Many changes are occurring in the cost structure of book publishing, which this study only hints at. It seems highly likely that arguably lower entry costs for publication of electronic books foster the expansion of electronic publishers that seems to be occurring. We have heard some evidence that the number of new print book publishers is growing, due in large part to lower entry costs through POD services, but that needs confirming. Certainly the cost of distribution and reaching markets is being redistributed, not only in electronic

publication but in traditional print publishing, as marketing support of publishers of all kinds shifts.

We have not seen a definitive study of the costs to the consumer of electronic books; the received wisdom is that costs of electronic books are lower, but that needs systematic documentation. Although this study uncovered numerous examples of authors who are distributing their work for free, the motivations are not clearly understood; surely readership and recognition is in the mix, but the extent to which authors see subsidized publication and free distribution as a path to recognition by a traditional print publisher is uncertain. Certainly well publicized examples exist, but how frequently they occur and how they affect the motivation for an author to self publish is, in our minds, still unclear.

74.5 Discovery, readership and reception and survival

In this arena, there are even more unanswered questions than in the areas above. In *The Long Tail*, Chris Anderson has extensively documented the shift from scarce resources in all media to abundance and the presence of many niche products and markets, and he has suggested that in an environment of abundance and small niche markets, the reader's ability to discover titles of potential interest to them is crucial (Anderson, 2006).

Channels to market play an important role in discovery, primarily through the creation of genre lists and search strategies. Increasingly, commercial publishers, authors, channel sites and niche interest sites of many varieties are experimenting with social forms of book discovery: reviews, recommendations, blogs, tagging, and other ways of sharing with like-minded people. In the romance arena, for example, reviews on popular sites like *Coffee Time Romance* or *Romance Junkies* cover print and electronic books in niche markets as well as mainstream ones.

An obvious question to ask in view of the increasingly abundant supply of books is whether they are finding readers. Anderson's argument, in terms of economics, is that a large number of small purchases, in the long tail, can be economically viable. Perhaps implied in that assertion is some restructuring or shifting of economic models of publishers and channels to market. Another question begging for hard data is how much electronic books, from whatever publishing source, are being purchased or borrowed, and read. Survival of electronic, and indeed POD books, is also a issue that raises more questions than answers, as in all areas of digital preservation. John B. Thompson (2004) documents the struggles of publishers to shift from the mindset that their chief asset is a published book in a warehouse to understanding their assets as electronic files saved and available in reformatable condition. And the debates around the preservation of e-journals in an environment where access has shifted from ownership to licensing under restricted conditions can be extended to electronic books, as can the on-going conversations of the long-term viability of today's electronic formats.

75. FUTURE DIRECTIONS

This study is our initial look at shifts in publishing practices in the age of the Internet and raises many more questions than it answers. Each of the observations above, and others, needs to be probed in greater detail. Data-driven studies of all sorts will be useful in documenting the emerging dimensions of shifts in book publishing in the 21st century.

76. REFERENCES

- [115] Adams, T. R., & Barker, N. (2001). A new model for the study of the book. In N. Barker (Ed.). *A Potencie of Life: Books in society* (pp. 5-43). London: British Library and Oak Knoll Press. First published in 1993 by the British Library.
- [116] Anderson, C. (2006). *The Long Tail*. Hyperion.
- [117] Choi, C. (2007). Got a manuscript? Publishing now a snap. Associated Press. Jan 2.
- [118] Hyatt, M. (2005) *The Death of the traditional book. Working Smart*.
<http://www.michaelhyatt.com/workingsmart>. Dec 11.
- [119] Keen, A. *Cult of the Amateur: How the Internet is Killing Culture*. 2006. Doubleday.
- [120] Thompson, J.B., (2005). *Books in the Digital Age*. Cambridge UK, Polity Press.

FULL PAPER SUBMISSION 33

Information Dissemination and Information Assurance in Vehicular Networks

AUTHOR(S):

Krishnamurthy, Prashant

ORGANIZATION(S):

University of Pittsburgh, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Vehicular networks aimed toward providing roadside services such as traffic alerts, estimated time to reach a destination, alternative routes, and in general improve the efficiency and safety on the road are emerging in both the United States and Europe. Information exchange in such networks occurs between vehicles (inter-vehicle communications) in an ad hoc manner and also with roadside base stations using so-called dedicated short range communication links. Research on technology related to vehicular networks is being conducted by many universities [JSAC07] and is being widely reported in the mainstream media as well [NYT07]. Vehicular networks are thus expected to become an important part of community networks of the future.

In this paper we will survey the different types of dissemination of information and the assurance of such information in vehicular networks. The paper will discuss the architecture of vehicular networks, classify different types of information exchange (such as opportunistic exchange of resources between vehicles [XU04], vehicle assisted data delivery [ZHAO07], cooperative downloading of information, [NAN05] etc.) and discuss the dynamics of the creation of sub-networks and group communications (e.g., [WU04]) (which can depend upon the density of vehicles on the road, speed of vehicles etc.). Then we discuss information assurance issues in vehicular networks (e.g., [RAYA06]) and survey the solutions proposed for ensuring authenticity/integrity of information, location privacy of vehicles, eviction of faulty or misbehaving vehicles from the information network (e.g., using reputation), etc. Where appropriate or suitable, we compare vehicular networks with other community/social networks.

References

- [JSAC07] IEEE Journal on Selected Areas in Communications, Special Issue on Vehicular Networks, October 2007.
- [NAN05] A. Nandan et al., "Co-operative downloading in vehicular ad-hoc wireless networks," Second IEEE Conference on Wireless On-demand Network Systems and Services, January 2005.
- [NYT07] Roy Furchgott, "Navigating With Feedback From Fellow Drivers," New York Times, October 18, 2007.
- [RAYA06] M. Raya, P. Papadimitratos, J-P. Hubaux, "Securing Vehicular Communications," IEEE Wireless Communications, October 2006.
- [WU04] H. Wu et al., "Analytical models for information propagation in vehicle-to-vehicle network," Proc. IEEE Vehicular Technology Conference, 2004.
- [XU04] B. Xu, A. Ouksel, O. Wolfson, "Opportunistic resource exchange in inter-vehicle ad-hoc networks," Proc. IEEE Int. Conf. on Mobile Data Management, pp. 4-12, 2004.
- [ZHAO07] J. Zhao and G. Cao, "ADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks," IEEE Trans. Vehicular Technology, Accepted for Publication.

FINAL (FULL) VERSION: (SEE THE FOLLOWING PAGE IF PAPER DOES NOT APPEAR BELOW)

Vehicular networks aimed toward providing roadside services such as traffic alerts, estimated time to reach a destination, alternative routes, and in general improve the efficiency and safety on the road are emerging in both the United States and Europe. Information exchange in such networks occurs between vehicles (inter-vehicle communications) in an ad hoc manner and also with roadside base stations using so-called dedicated short range communication links. Research on technology related to vehicular networks is being conducted by many universities [JSAC07] and is being widely reported in the mainstream media as well [NYT07]. Vehicular networks are thus expected to become an important part of community networks of the future.

In this paper we will survey the different types of dissemination of information and the assurance of such information in vehicular networks. The paper will discuss the architecture of vehicular networks, classify different types of information exchange (such as opportunistic exchange of resources between vehicles [XU04], vehicle assisted data delivery [ZHAO07], cooperative downloading of information, [NAN05]] etc.) and discuss the dynamics of the creation of sub-networks and group communications (e.g., [WU04]) (which can depend upon the density of vehicles on the road, speed of vehicles etc.). Then we discuss information assurance issues in vehicular networks (e.g., [RAYA06]) and survey the solutions proposed for ensuring authenticity/integrity of information, location privacy of vehicles, eviction of faulty or misbehaving vehicles from the information network (e.g., using reputation), etc. Where appropriate or suitable, we compare vehicular networks with other community/social networks.

References

- [JSAC07] IEEE Journal on Selected Areas in Communications, Special Issue on Vehicular Networks, October 2007.
- [NAN05] A. Nandan et al., "Co-operative downloading in vehicular ad-hoc wireless networks," Second IEEE Conference on Wireless On-demand Network Systems and Services, January 2005.
- [NYT07] Roy Furchgott, "Navigating With Feedback From Fellow Drivers," New York Times, October 18, 2007.
- [RAYA06] M. Raya, P. Papadimitratos, J-P. Hubaux, "Securing Vehicular Communications," IEEE Wireless Communications, October 2006.
- [WU04] H. Wu et al., "Analytical models for information propagation in vehicle-to-vehicle network," Proc. IEEE Vehicular Technology Conference, 2004.
- [XU04] B. Xu, A.Ouksel, O. Wolfson, "Opportunistic resource exchange in inter-vehicle ad-hoc networks," Proc. IEEE Int. Conf. on Mobile Data Management, pp. 4-12, 2004.
- [ZHAO07] J. Zhao and G. Cao, "ADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks," IEEE Trans. Vehicular Technology, Accepted for Publication.

Posters

(30 Accepted Poster Submissions)

Session Chair: Jonathan Furner

POSTER SUBMISSION 1:

Towards a Model of Determinants of Web Services Platform Adoption by Complementers

AUTHOR(S):

Rubleske, Joseph B.

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The recent surge of interest in web services has called attention to the increasingly intense competition between owners of the platforms on which these services run. Given that widely adopted operating systems and middleware platforms have yielded sizable economic returns for their owners, many web services platform owners are aggressively pursuing strategies that can give them a competitive advantage and, it is hoped, similarly sizable returns.

A review of the broader literature on software platform competition reveals widespread acceptance of network effect theory as an explanatory framework. Network effect theory posits that the value of a software platform to a potential user is associated positively with the number of existing users of the platform (who generate direct network effects) and the number of developers of complementary software applications (who generate indirect network effects) (see, e.g., Katz and Shapiro, 1986; Zhu et al., 2006). Users realize direct network effects when, for example, they share compatible files with other users (Gao and Iyer, 2006; Lin and Kulatilaka, 2006) or participate in 'trading communities' (Zhu et al., 2006). Indirect network effects are realized through the availability of useful, innovative and compatible software applications (Lin and Kulatilaka, 2006). Users of widely adopted software platforms also gain value from the reduced likelihood of being "stranded with a failed and unsupported platform" and consequent switching costs (Gallaughar and Wang, 2002, p. 306).

In the presence of network effects, then, software platform owners pursue strategies that will secure them an 'installed base' of users and complementers that is sufficiently large to attract more and more new users (Shapiro and Varian, 1998; Suarez, 2005). While one set of strategies is aimed at promoting adoption by new users, another set emphasizes the value generated for users by indirect network effects and aims instead at promoting adoption by complementers. (This distinction reflects the idea that platform markets are two-sided, with (end) users populating one side and complementers populating the other.)

There appears to be considerably more research on strategies for increasing user adoption (see Gallaughar and Wang (2002), von Westarp (2003) and Zhu and Iansiti (2007) for reviews) than on complements adoption strategies. Nonetheless, three studies of the latter merit mentioning here. First, in their study of the U.S. video game industry from 1976 to 2002, Venkatraman and Lee (2003) find that platform dominance (i.e., largest installed base), together with complementers' path dependency and level of experience with platform architecture, largely determine platform adoption by complementers. Second, in his investigation of how software platform owners maintain a balance between "adoption and appropriation," West (2003) concludes that software platform owners who disclose some proprietary code will attract more complements (thereby fostering innovation), but cautions against disclosing any code that confers a competitive advantage. Finally, Cusumano and Gawer's (2002) landmark study of Intel's platform management strategies culminated in the endorsement of four 'levers' for platform leadership, with one of these levers aimed at managing relations with 'external complementers'. Specific strategies include building a consensus on technical specifications and standards, handling potential conflicts of interest and letting complementers keep any intellectual property they develop on the platform. Both West (2003) and Cusumano and Gawer (2002) also underscore the importance of providing complementers with an interface to connect to the platform. Beyond West's (2003, p. 1260) suggestion that software platform owners "create and

evolve application programming interfaces (APIs),” though, the varied ways in which these APIs might influence a complements’ choice to adopt have not been sufficiently explored by these or other authors.

The research-in-progress described in the following section aims to bolster the somewhat scant literature on software platform adoption by complements. More specifically, the proceeding research design outlines a proposed investigation of the determinants of complements’ adoption of geo-mapping web services platforms. The reasons for including independent variables are discussed, and some methodological details are introduced. The paper concludes with a brief discussion of anticipated outcomes of the study.

Research Design

This section proposes a working model for investigating the determinants of complements’ adoption of web services platforms. A web services platform is defined in this paper as a software program that makes web services (i.e., interoperable machine-to-machine interactions) available to other software programs through application programming interfaces, or APIs (Evans et al., 2006). APIs are an essential component of web services platforms, as they specify – through a set of conventions – how developers can request and receive web services. For example, Flickr.com (which stores and supports the sharing of digital photos) offers an API that will return to successful ‘calls’ to its database a list of all photos uploaded to Flickr during a certain period of time, all the comments about certain photos, photo metadata and much more. Developers can also upload batches of photos and add tags to them through Flickr’s API. (As of this writing Flickr’s API consists of 59 other methods like these.) In sum, APIs instruct developers on how to configure their programs so they can access data from the web service that provides the API.

The web site Programmableweb.com maintains an extensive directory of web services platforms. As of this writing, this directory includes nearly 550 APIs for more than 400 web services platforms. (Some of the larger platform owners such as AOL, Google, Microsoft and Yahoo! offer many APIs for their web services platform.) Each of these APIs is organized into one of 53 functional or topical categories (e.g., advertising, chat, financial, mapping, news, search). Programmableweb.com also collects up-to-date information about each of these APIs, including the number of mashups that have made use of the API. Just as a musical mashup is a recording that combines two or more musical sources, a web mashup is an application that combines data and/or functionality from two or more web services.

Web mashups are relevant to our study because they generate complementary assets for the web services platforms on which they are built. Indeed, given that the manufacture of an exhaustive list of web services platforms is not a feasible task, the web mashups listed in Programmableweb.com serve in our study as a surrogate measure of web services adoption by complements. Additional information from Programmableweb.com about each API – such as language protocols, signup requirements, usage fees and more – is also integral to our study, as some of this information will be operationalized as determinants of complements’ adoption.

Rather than proceeding at the outset to analyze all 550 APIs, a pilot study of the 48 APIs comprising the geo-mapping category is proposed. We believe that the geo-mapping category may be best suited for our pilot test of the working model because

- It includes more APIs (48) than any other category;
- Two-thirds (1,589) of all mashups listed on Programmableweb.com make use of at least one geo-mapping API;
- Google’s platform dominates (1,194 of the 1,589 geo-mapping mashups make use of its Google Maps API), but the Microsoft and Yahoo! platforms are vying for second place through their respective APIs;
- The APIs of three lesser-known platforms (geocoder, GeoNames and Multimaps) have each been used in 10 or more mashups, suggesting that a large installed base of users may not be the sole determinant of complements’ adoption; and
- Twenty-four (24) APIs have each been used in somewhere between one and nine mashups, suggesting that the geo-mapping category exhibits ‘power law’ scaling with the distinguishing ‘long tail’.

The working model’s potential determinants of complements’ adoption (see Figure 1) are derived in part from reviews of Cusumano and Gawer (2002) and West (2003) and of archived feeds from several blogs written by software engineers, mashup developers, industry analysts and internet entrepreneurs. Potential determinants that

can only be evaluated through surveys and interviews (i.e., not from publicly available sources) were set aside to be examined in a future, later-phase study. For example, the idea that users adopt software platforms which they expect to become dominant (Shapiro and Varian, 1998; Zhu and Iansiti, 2007) may apply to complementers as well, but it cannot be tested without surveying them. Figure 1 illustrates why these potential determinants are relevant and identifies the likely data sources for each one.

Anticipated Outcomes

While most studies of user adoption of software platforms have confirmed the importance of a large installed base, many of these studies have qualified their findings by noting that other factors can also be very influential under certain conditions. Given the prevailing rhetoric (in the web services industry) about ‘monetization opportunities’ for complementers, I would not be surprised to find that economic factors largely drive complementer adoption, nor would I be surprised to discover other relevant multivariate relationships not mentioned in the software platform adoption literature – particularly in light of the fact that the proposed dataset has not been explored in any peer-reviewed studies. In conclusion, I believe that the findings of this research-in-progress will have implications for network effect theory (and possibly for the development of new concepts) and for web services platform owners desiring more effective strategies for generating complementary assets.

Figure 1: A Working Model of Potential Determinants of Complementer Adoption

Independent Variables	Justification	Source(s) for Data
1.Larger installed base (+)	Network effect theory	Week-by-week adoption rates (since API release) compared to intalled base statistics (including all users)
2.Greater opportunities to generate revenue (i.e., monetization opportunities) (+)	Complementers want to maximize economic returns	Programmableweb.com, API and mashup web sites, API and mashup blogs, API and mashup forums
3.Lower administrative barriers to implementation (+)	Complementers want to be able to work with an API right away	Programmableweb.com ('signup requirements')
4.More functional allowances (+)	Complementers prefer APIs that offer more functionality, not less	Programmableweb.com, API and mashup web sites, API and mashup blogs, API and mashup forums, first-hand use of APIs, etc.
5.Greater use of open standards (+)	Complementers prefer APIs based on W3C-approved standards, not proprietary ones	Programmableweb.com ('protocols' and 'data formats')
6.More extensive supportive content (+)	Complementers prefer APIs that offer more supportive content, not less	API web sites, API blogs, API forums, etc.

References

1. Cusumano, M.A. and Gawer, A. "The elements of platform leadership," MIT Sloan Management Review (43:3), 2002, pp. 51-58.
2. Evans, D.S., Hagiu, A. and Schmalensee, R. Invisible Engines: How Software Platforms Drive Innovation and Transform Industries, MIT Press, Cambridge, MA, 2006.
3. Gallagher, J.M. and Wang, Y. "Understanding network effects in software markets: evidence from web server pricing," MIS Quarterly (26:4), 2002, pp. 303-327.
4. Gao, L.S. and Iyer, B. "Analyzing complementarities using software stacks for software industry acquisitions," Journal of Management Information Systems (23:2), 2006, pp. 119-147.
5. Katz, M. and Shapiro, C. "Technology adoption in the presence of network externalities," Journal of Political Economy (94:4), 1986, pp. 822-841.
6. Lin, L. and Kulatilaka, N. "Network effects and technology licensing with fixed fee, royalty, and hybrid contracts," Journal of Management Information Systems (23:2), 2006, pp. 91-118.
7. Shapiro, C. and Varian, H.R. Information Rules: A Strategic Guide to the Network Economy, Harvard Business School Press, Boston, MA, 1998.
8. Suarez, F.F. "Network effects revisited: the role of strong ties in technology selection," Academy of Management Journal (48:4), 2005, pp. 710-720.
9. von Westarp, F. Modeling Software Markets: Empirical Analysis, Network Simulations and Marketing Implications, Physica-Verlag, Heidelberg, Germany, 2003.
10. West, J. "How open is open enough? Melding proprietary and open source platform strategies," Research Policy (32), 2003, pp. 1259-1285.
11. Zhu, F. and Iansiti, M. "Dynamics of platform competition: exploring the role of installed base, network quality and consumer expectations," Harvard Business School Working Paper, Boston, MA, 9 October 2007.
12. Zhu, K., Kraemer, K.L., Gurbaxani, V. and Xu, S.X. "Migration to open-standard interorganizational systems: network effects, switching costs, and path dependency," MIS Quarterly (30), 2006, pp. 515-539.

POSTER SUBMISSION 2:

Towards a data and workflow collaboratory for research on free and open source software and its development

AUTHOR(S):

Howison, James
Crowston, Kevin
Conklin, Megan

ORGANIZATION(S):

Syracuse University, United States of America

Elon University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Towards a data and workflow collaboratory for research on free and
open source software and its development
A poster submission for iConference 2008

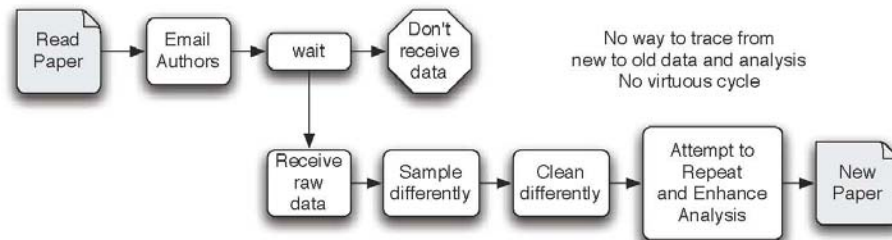
(Author details redacted)

29 October 2007

FLOSSmole is a collaborative data repository which collects and provides data for research on Free/Libre Open Source Software (FLOSS) and its development by online, distributed teams. The data is used by a research community that studies diverse questions from the evolution of software to how these groups make decisions, use various media and manage change over time (Scacchi, 2007). This multi-disciplinary research community includes researchers both inside and outside iSchools, from many disciplines including software engineering, organizational studies, information systems and sociology, as well as corporate researchers.

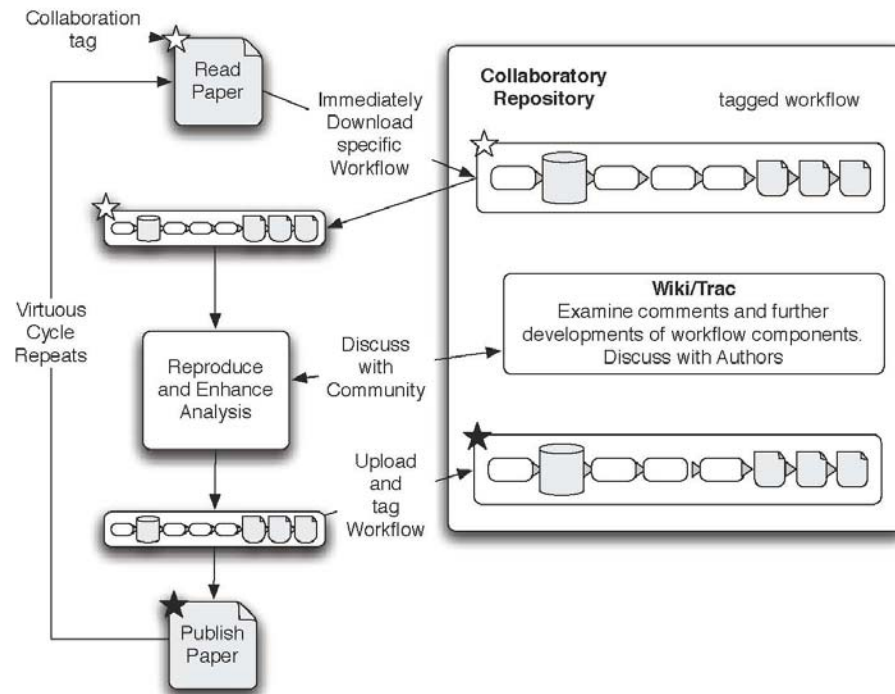
Since May 2007, we¹ have been working to development this repository into an e-Social Science infrastructure capable of supporting, storing and publishing not only data but analyses organized into scientific workflows, as envisioned in the NSF reports on Cyberinfrastructure (Atkins, 2003; NSF Cyberinfrastructure Council, 2007). Further we are developing a pre-print repository which will enable bidirectional links between the data, workflows and published papers. The goal of the project is to facilitate collaboration to improve the reproducibility and consistency of research, collaboratively building a body of cumulative knowledge (Borgman, 2007; Finholt and Olson, 1997).

¹References to our FLOSS research have been excluded for review

Figure 1: *Current research practice has significant difficulties*

The current social science research on FLOSS has relied on several different kinds of scientific evidence, including the archives created by the FLOSS developers, versioned code repositories, mailing list messages and bug and issue tracking repositories. FLOSS teams retain and make public archives of many of their activities as by-products of their open technology-supported collaboration. However, the easy availability of primary data provides a misleading picture of ease of conducting research on FLOSS. Precisely because data are by-products, they are not in a form that is useful for researchers. Instead potentially useful data is locked up in HTML pages, CVS log files, or text-only mailing list archives. Furthermore because FLOSS projects are hosted in a variety of “forges” (of which Sourceforge is but the largest of dozens) or individual websites, these problems are multiplied. Finally different projects use similar services in quite different ways, leading to inappropriate comparisons. Research projects expend significant energy collecting and re-structuring these archives for their research. This process is repetitive, wasteful and produces inconsistent results.

The situation with analyses and workflows is even more inconsistent and repetitive, with different research groups performing similar sampling, data cleaning and analyses steps but unable to share or to build on each others’ work. Our project is learning from e-Science infrastructure in the natural sciences, which provide shared databases for data such as genome sequences and astronomical observations, often hosted under the purview of the NSF’s Ter-

Figure 2: *Virtuous cycles enabled by research workflow sharing*

raGrid project. They have developed graphical desktop software, such as Taverna² and Kepler³, to link together these data with analytical transformations which can be made available as downloaded components or web-service gateways to distributed computational resources ('grid computing'). Increasingly it is also possible to include commercial services, such as Yahoo's Pipes or Amazon's storage and computational resources.

Crucially the workflows can be stored and shared (as XML files) and an online service is provided to 'stamp' a particular workflow configuration with a unique identifier for later

²<http://taverna.sourceforge.net/>

³<http://kepler-project.org/>

reference. Each ‘run’ of the workflow produces a record of the transformations undertaken, preserving the provenance of the analytical results. It is hoped that our project will facilitate the use and sharing of such workflows, allowing FLOSS researchers to re-use appropriate and tested components, freeing them to concentrate their efforts on novel contributions (‘return from scripting to science’). Such documentation and re-use is valuable both between research groups but also within individual research groups over time, as collaborators and students come and go.

As useful as these natural science systems are, their application to social science requires significant adaptation (Berman and Brady, 2005), which our poster presentation will highlight for the benefit of the multi-disciplinary iSchool audience. For example, the FLOSS research community commonly draws on content analytic techniques to investigate textual archives, such as mailing lists. This content analysis can be both positivist (abstracting observations, often by producing counts or sequences of observed concepts) or interpretative (focused on accessing the meaning for participants). These ‘marked-up’ texts, as well as the content analytic schemes developed, are a resource which could be valuable to other researchers. However it is vital that the ‘workflow’ of their production is also clearly described together with intermediate artifacts like inter-coder reliability statistics. Our project is exploring appropriate ontologies, like those of the Data Documentation Initiative⁴, and will be able to present preliminary results in February.

Finally, publications refer to data and measures using inexact and conflicting descriptions, complicated by the fact that researchers often draw on the same raw data to operationalize different concepts. It is hoped that our research will create a controlled namespace, allowing researchers to move both from publications to specific data and workflows, and to see quickly how particular data or analysis techniques have been used in published papers.

In order to demonstrate the value of such techniques, and to kick-start their use in this research community, our project will engage in digital curation, selecting six important papers

⁴<http://www.ddalliance.org/>

which will be replicated using the adapted e-Science techniques. The data selections and analysis components so created will be made available as base infrastructure and introduced at an NSF-funded FLOSS research community workshop to be held just prior to the iSchool conference.

In summary, the proposed poster will stimulate discussion about multi-disciplinary scientific ecosystems, requirements for adapting Cyberinfrastructure and e-Science techniques for the social sciences and provide a template for those in the iSchools community seeking to enhance collaboration and quality in their research communities.

Bibliography

- Atkins, D. (2003). Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure.
- Berman, F. and Brady, H. (2005). Final report: NSF SBE-CISE workshop on cyberinfrastructure and the social sciences. Available at www.sdsc.edu/sbe/.
- Borgman, C. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. MIT Press.
- Finholt, T. A. and Olson, G. M. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, 8(1):28-36.
- NSF Cyberinfrastructure Council (2007). Cyberinfrastructure vision for 21st century discovery. NSF Report 0728.
- Scacchi, W. (2007). Free and open source software development: Recent research results and methods. In Zelkowitz, M., editor, *Advances in Computers*, volume 69. Elsevier Press.

POSTER SUBMISSION 3:

Behavioral Parameters of Trustworthiness for Countering Insider Threats

AUTHOR(S):

Ho, Shuyuan Mary

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The Third Annual iConference 2008

Research-in-Progress for Poster

Behavioral Parameters of Trustworthiness for Countering Insider Threats

(abstract)

Introduction

Since Robert Hanssen, a US counterintelligence agent, started spying and gave away highly classified national security documentary materials to KGB¹/SVR² in Soviet Union / Russia in 1970, the case of a betrayal of trust by a trusted, high-ranked insider was established (FBI National Press Office, 2001). This case portrays that not only the trust level of a key person with high-level security clearance could be altered, but the danger s/he brings to corporate security is maximized as s/he knows what and where the critical corporate resources are. In the *Insider Threat Study by CERT (2004-2005)*, US DoD³, DHS⁴, & Secret Service investigated various insider threat cases and discovered that embedded in a mesh of communications, a person given high social power but with insufficient trustworthiness can create a single point of trust failure (Randazzo, et al., 2004; Keeney, et al., 2005). Thus, “insider threats” as an organizational problem gap is defined as executives or someone with authorized access, high social power and holding a critical job position, who is capable of inflicting high impact damage including psychological, managerial, or physical level, within an organization.

Conceptualization**Conceptualizing Trustworthiness**

Figure 1 depicts the conceptual framework of my study into trustworthiness. It utilizes a multi-level analysis, including mixed “lenses” of organizational and individual norms.

¹ KGB (transliteration of “КГБ”) is the Russian abbreviation for Committee for State Security (Комитет Государственной Безопасности).

² SVR is the Russian abbreviation for Foreign Intelligence Service (Служба Внешней Разведки), which is Russia’s primary external intelligence agency.

³ US DOD stands for the U.S. Department of Defense.

⁴ DHS stands for the Department of Homeland Security.

Shuyuan Mary Ho

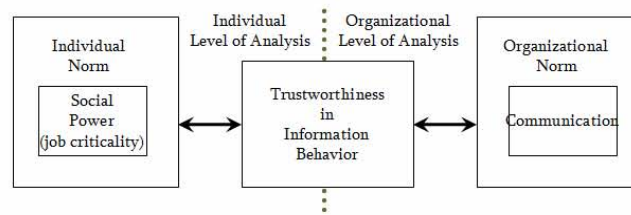


Figure 1: Conceptual Framework

Julian Rotter (1971) believes that “trust and trustworthiness are closely related.” Rotter (1980) defines trust as a generalized expectancy - held by an individual or a group - that the communications of another (individual or group) can be relied upon (Rotter, 1967, p. 652) regardless of situational specificity (Rotter, 1980). Hardin (1996), on the other hand, differentiates trustworthiness from trust. He believes that trustworthiness is “a moralized account of trust” (Hardin, 1996, p. 28). For example, a criminal (A) can trust a criminal-partner (B) to conduct a joint crime and there is no moral or ethical notion involved. However, it might take much more complexity in decisions for A to let B handle A’s financial accounts because A may not find B trustworthy (Hardin, 1996, p. 29). In this light, I define an employee’s “trustworthiness” as the generalized expectancy, a subjective probability, toward a target’s degree of correspondence between communicated intentions and behavioral outcomes that are observed and evaluated over time. In other words, the degree of correspondence between the target’s words and actions remain reliable, ethical and consistent, which its degree of fluctuation does not exceed observer’s expectations over time. To be able to tell if an employee is trustworthy is thus determined by the subjective perceptions of individuals in his/her social network who have direct business functional connections, and thus the opportunity to repeatedly observe correspondence of communication and behavior.

Attribution Theory Review

Attribution Theory intends to understand how people attribute (or assign) causes to another’s behavior (Heider, 1944 & 1958). It’s a cognitive perception. Attribution theory dichotomizes behavioral causes to both internal and external. If the causes of behavior are attributed to the person, it is called the internal attribution. Such causality of behavior is both internal and dispositional. If the causes of behavior are attributed to the situation, it is considered external attribution. The basic observational “setting” contains three major variables: the observer, the target and the situation. Perception can vary if the observations (or interpretation) are from different observers, or if the target being studied is different, or if the situation is different (Heider, 1958; Weiner, 2006, p. xvii).

Shuyuan Mary Ho

Kelley (1973) suggests a causal attribution theory where an observer has multiple sources of relevant information, and is likely to infer and detect the causes of observed behavior. The observed behavior can be interpreted and perceived in a single observation – or through multiple observations over time. Kelley (1973) suggests three principles, distinctiveness, consensus and consistency, be applied in these multiple observations.

Research Question

I am interested in studying the basic mechanisms for detecting changes in the trustworthiness of an individual who holds a key position in an organization, by observing overt behavior – including communication behavior – over time (see, for example, Steinke, 1975). Since Steinke suggests that it is possible to detect cheating behavior without directly observing the individual, my overarching question with regards to insider threat phenomenon is: *Why are the clues to a critical worker's future behavior so difficult to detect by members of a community?* Specifically, my research questions can be rephrased to three continuous sub-questions: With regards to personnel in authority positions, is it possible to detect changes in trustworthiness from subjective reflections and indirect perceptions of his/her social networks (peers, subordinates or associates)? If yes, what are the basic mechanisms of detecting changes in an individual's trustworthiness level? Would it then be possible and reliable to predict the potential for insider threats in advance using these subjective reflections?

Methodology

I intend to design experiments to answer my research questions. The perception given by observers can vary depending on the interpretations of different observers, the target, and different situational settings (Heider, 1958; Weiner, 2006). The attribution of the target's (A's) behavior by observers (B's) is believed and determined by B's judgment that A intentionally or unintentionally (Heider, 1958) behaves in a way that the cause of behavior is attributable to either external (situational) causality or internal (dispositional) causality.

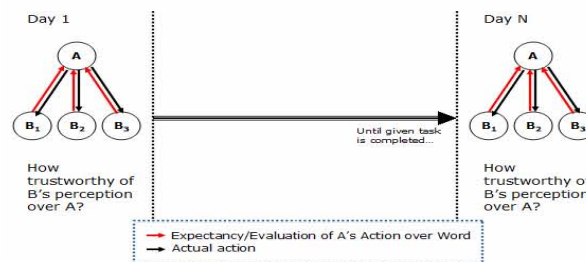


Figure 2: Illustration of Experimental Observation Over Time

Shuyuan Mary Ho

The principle of distinctiveness shall be applied to the target's behavior. In another words, the target's behavioral change has to be noticeable for others to perceive it. In order to eliminate participant bias, consensus among observers should be obtained whenever possible. The consistency between the target's words and actions shall be evaluated by these observers (Kelley, 1973) over time (Day 1 through Day N) until a given task is completed (Figure 2).

Operationalization Rationale

Consistency between the target's (A's) words and actions is an important indicator to properly measure generalized expectancy of this target's social communication to and from his/her peers. The real focus of the study is to see how other people characterize the attributes of this leader during slightly suspicious behavior. Some challenging considerations and innovative ideas are utilized in the design of this experiment. 1.) Scenario encapsulation by a shielded title so that the unwanted participant acting bias can be filtered out. 2.) Manipulating the target through forcefully creating a dishonesty gap (Figure 4) between the interests of the leader and of the team players. 3.) A real-world case is simulated through feeding information to the participants in both public and private settings. 4.) The target is empowered through direct appointment. 5.) The art of "fishing dynamics" is introduced in the game scene while an authoritative figure is manipulating the target with an ethical dilemma. 6.) Team involvement is quickened through fun and competitive brain teasers. 7.) Experiment is designed to be flexible and can be repeated through "virtual asynchronous contest."

Experiment Design

This experiment design is called the "leader's dilemma" game. This experiment implements the concept of a "virtual asynchronous contest," and is designed to recruit one real team, and share fictitious scores of three other teams (Figure 3).

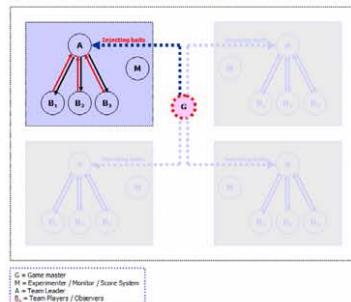


Figure 3: Controlled Room Design of Pilot Study

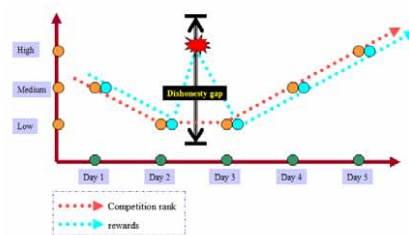


Figure 4: Logics for Virtually Controlled Contest

The bright blue area is the actual team recruited - the gray areas represent fictitious teams. I plan to recruit 4 or 5 participants each experiment. One participant will be appointed as the leader (A). The remaining participants will be team players (B_n). The experimenter (M) will

Shuyuan Mary Ho

be monitoring the game play and collecting observations. M's role in this controlled environment will be that of a positive judge. The game-master (G) is in an authoritative position, but has a slightly negative role. G has knowledge of the competition and has the power to award the winning team. G's job is to inject the bait (in the form of micro-currency) to entrap A. While A is entrapped, B_n perceives A's behavior. The perceptions from B_n will be collected in various ways: 1) daily survey from B_n , 2) participant observation from M, and 3) the individual semi-structured interview at the end of the experiment of both A and B_n by the experimenter, M. This virtual contest reaches its climax when a dishonesty gap is forcefully created by feeding the bait to the team leader alone (Figure 4).

Study Contribution and Limitation

In traditional information security strategy, in order to effectively quarantine polymorphic virus codes, it has been necessary to study how codes change. Likewise, we may be able to detect suspicious behavior and counter insider threats by studying how human behavior changes and how his/her trustworthiness is altered through indirect perception of others within an organization. This study provides basic mechanisms, clues, and early warning signs to investigate and detect fluctuating personnel trustworthiness; however it is important to note that these by no means possess full assurance to convict crimes. While these basic mechanisms could be adopted in cognitive modeling of detection systems, human intervention is still necessary in this loop.

Shuyuan Mary Ho

References

- FBI National Press Office. (2001). *Federal Beureau of Investigation Story: Robert Philip Hanssen Espionage Case*. Released on Feb 20, 2001. Obtained from <http://www.fbi.gov/libref/historic/famcases/hanssen/hanssen.htm>.
- Hardin, R. (1996). Trustworthiness. *Ethics*, Vol. 107, No. 1. (Oct., 1996), pp. 26-42.
- Heider, F. (1944). *Social perception and phenomenal causality*, *Psychological Review*, 51, 358-374.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons. ISBN: 0-471-36833-4.
- Ho, S.M. (2008). *Towards a Deeper Understanding of Personnel Anomaly Detection*. Encyclopedia of Cyber Warfare and Cyber Terrorism, 2008 IGI Global Publications, Hershey, PA.
- Keeney, M., Kowalski, E., Cappelli, D., Moore, A., Shimeall, T., and Rogers, S. (2005). "Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors." National Threat Assessment Center, U.S. Secret Service, and CERT® Coordination Center/Software Engineering Institute, Carnegie Mellon, May 2005, pp.21-34. Obtained from <http://www.cert.org/archive/pdf/insidercross051105.pdf> on April 10, 2007.
- Kelley, H.H. (1973). The Process of Causal Attribution, *American Psychologist*, Feb 1973, 107-128. Obtained from http://faculty.babson.edu/krollag/org_site/soc_psych/kelly_attrib.html on July 5th, 2007.
- Randazzo, M.R., Keeney, M., Kowalski, E., Cappelli, D., and Moore, A. (2004). Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector. National Threat Assessment Center, U.S. Secret Service, and CERT® Coordination Center/Software Engineering Institute, Carnegie Mellon, August 2004. http://www.secretservice.gov/ntac/its_report_040820.pdf.
- Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35 (4), 651-665.
- Rotter, J.B. and Stein, D.K. (1971). Public Attitudes Toward the Trustworthiness, Competence, and Altruism of Twenty Selected Occupations. *Journal of Applied Social Psychology*, Dec 1971, 1(4), 334-343.
- Rotter, J.B. (1980). Interpersonal Trust, Trustworthiness, and Gullibility. *American Psychologist*, Jan 1980, 35(1), 1-7.
- Steinke, G.D. (1975). The prediction of untrustworthy behavior and the Interpersonal Trust Scale. Unpublished doctoral dissertation, University of Connecticut, 1975.
- Weiner, B. (2006). *Social Motivation, Justice, and the Moral Emotions: An Attributional Approach*. Lawrence Erlbaum Associates, inc., Mahwah, New Jersey.

POSTER SUBMISSION 4:

ChemXSeer: An eScience Web Search Engine and Repository for Chemistry

AUTHOR(S):

Giles, C. Lee
Mitra, Prasenjit
Muller, Karl
Kubicki, James
Garrison, Barbara
Wang, James Z.
Sun, Bingjun
Liu, Ying
Tan, Qingzhao
Bolelli, Levent
Lu, Xiaonan
Jaiswal, Anuj
Bai, Kun
Councill, Isaac
Brouwer, William
Fernandez, Juan
Bandstra, Joel

ORGANIZATION(S):

Pennsylvania State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

1. Introduction

E-science or cyberinfrastructure have become crucial for scientific progress and open source systems have greatly facilitated design and implementation. In chemistry, the growth of data has been explosive and timely and effective information and data access is critical [Atkins 2003, Hey 2006]. Many have argued that cyberinfrastructures for science are domain sensitive [Snow 2006] and many have been proposed. We have proposed and are developing the ChemXSeer architecture, a portal for academic researchers in environmental chemistry, which integrates the scientific literature with experimental, analytical and simulation datasets. ChemXSeer will be comprised of information crawled from the web, manual submission of scientific documents and user submitted datasets, as well as scientific documents and metadata provided by major publishers. Information crawled by ChemXSeer from the web and user submitted data will be publicly accessible whereas access to publisher resources can be provided by linking

to their respective sites. Thus, instead of being a fully open search engine and repository, ChemXSeer will be a hybrid one, limiting access to some resources.

2. Chemistry and Cyberinfrastructure

Chemical research is becoming increasingly collaborative in scope and approach. For example, within the Penn State Center for Environmental Kinetics Analysis (CEKA), researchers are taking a multidisciplinary approach to linking kinetic information in environmental chemistry across spatial and temporal scales (see www.ceka.psu.edu for more information). A main goal of such research is to integrate experimental, analytical, and simulation results performed on systems

from molecular to field scales in order to better approximate the complex physical, chemical, and biological interactions controlling the fate and transport of contaminants. Researchers in this area have realized for some time that new scientific questions can be generated when users have access to a broad spectrum of related results. Subsequently, as connections are made among field observations, experimental kinetic results, spectroscopic analyses, and model predictions, gaps in the information web will become apparent. Approaches to

filling these gaps can then be addressed by the collaborative team. In addition, when inconsistencies among results are identified through use of databases and

collaboration tools, determining the reasons for these differences can motivate and guide researchers to formulate experiments to eliminate discrepancies or make new discoveries. In short, data collection and synthesis will lead to better science and improved education of scientists.

Academic chemistry is now focusing on cyberinfrastructure in many areas, with a focus on building tools that are open source and shared within their community. There is much interest in chemistry web services [Coles 2005], chemical markup languages [Murray-Rust 2001], chemical ontologies [ChEBI], chemical literature indexing and search [PubChem], data repositories [Fletcher 1996, Afeefy 2005], chemistry notebooks [Hughes 2004], and others. As part of this wave of development, the ChemXSeer system will focus on the needs of domain scientists in chemistry, especially the interdisciplinary needs that arise when addressing complex kinetics synthesis problems.

3. The ChemXSeer System

ChemXSeer intends to offer unique aspects of search not yet present in other scientific search services. We have developed or are developing algorithms for the

extraction of tables, figures, equations and formulae from scientific documents enabling users to search on those fields. ChemXSeer intends to or currently does provide the following search features:

- Full text search
- Author, affiliation, title and venue search
- Figure search
- Table search
- Formulae search
- Citation and acknowledgement search
- Citation linking and statistics

A beta version of ChemXSeer is online and working.

4. Data Search

For dataset search, we are developing tools that automatically annotate published data representations such as figures and that permit researchers to annotate

their datasets by providing both document-level and attribute-level metadata in OAI-PMH format [Jaiswal 2006]. This level of data annotation provides us with the

opportunity of searching data more effectively both at the attribute and semantic levels, browsing datasets and linking to existing scientific literature and other datasets in our and other repositories. The data search feature is now available as a beta in ChemXSeer

5. Chemical Formula Search

Current search engines do not recognize chemical entities (chemical names and formulae). A scientist, who seeks to search for information related to chemical

molecules from text documents cannot do so in any meaningful way except performing exact keyword search. We show how a chemical-entity-aware search engine can be designed and built and demonstrate empirically that it improves the relevance of the returned results for search queries involving chemical entities. Our search engine first extracts chemical entities from text, performs a novel indexing suitable for chemical formulae and names, and supports different query models that a scientist may require. We develop a model of hierarchical conditional random fields (HCRFs) for entity tagging that considers long-term dependencies at the higher (hierarchical) levels. Experiments and examples show that our approach

performs reasonably well [Sun 2007]. A beta of formula search is now available in ChemXSeer.

6. Table Search

Tables are ubiquitous in scientific documents where they are widely used to present experimental results or statistical data in a condensed fashion. However, current search engines do not support table search. Scientists and scholars often extract data from tables in documents by hand. We have created a table search engine (TableSeer) that eliminates that burden enabling users to automatically

search for and examine tables. TableSeer crawls scientific documents, detects tables from documents, extracts and indexes tables metadata, and enables end-users to search for tables. We also propose an extensive set of edium-independent metadata for tables because there exists no tablemarkup language and universal metadata

specification for tables that scientists and other users can adopt for representing table information in documents. We demonstrate the value of our specialized table-search engine by empirical studies using scientific documents [Liu 2006, 2007]. A beta of TableSeer is now available in ChemXSeer.

7. Figure Search

Like tables, much important information in documents resides in figures. We are interested in all of figure search. However, our current focus is on searching for and

extracting data from plots. Two-dimensional (2-D) plots in digital documents contain important information. Often, the results of scientific experiments and performance of businesses are summarized using plots. Although 2-D plots are easily understood by users, current search engines rarely utilize the information contained in the plots to enhance the results returned in response to queries. We have developed an automated algorithm for extracting information from line curves in

2-D plots. Experimental evaluation has demonstrated that our method can produce results suitable for real world use [Lu 2006, Lu 2007]. We intend to make a beta for plot search available in ChemXSeer.

7. Search Integration

Integration of search results is a nontrivial issue. We propose to investigate new methods for integrating chemical search based on molecular centric views permitting users to have easy access to all related information [Bolelli 2007]. We also intend to link search results to other available databases.

8. Conclusions

ChemXSeer is an ongoing cyberinfrastructure project in chemistry. By offering chemists automatic access to information and data not previously available, ChemXSeer will change the way chemists do research and will significantly enhance scientific productivity and contribute to new discoveries.

9. Acknowledgements

The ChemXSeer cyberinfrastructure research and project is supported in part by a grant from NSF Chemistry.

POSTER SUBMISSION 5:

Support Sense-Making with Tools for Structuring a Conceptual Space

AUTHOR(S):

Zhang, Pengyi

ORGANIZATION(S):

University of Maryland - College Park, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

- Think-aloud protocols to learn about users' evolving progress of the sense-making process;
- Intermediate representations and final documents produced by the tasks;
- Search and use activity logs automatically recorded by the system.

Results from the user studies will be used to refine and improve the functions of the tool until the delivery of the final system.

5. Conclusions and Implications

Sense-making tools are useful in many settings. They can be used for knowledge management by individuals or the sharing of knowledge in small groups for collective sense-making. Different formats allow differences in representation and may allow users with a range of cognitive styles to share knowledge that is otherwise hard to share.

Helping users retrieve the right information is only half the battle; assisting users with making sense of what they found is the next frontier in information system design. This study will contribute to our understanding of sense-making processes, and thereby give a better foundation for system design.

Reference:

- Dervin, B. (1998). Sense-Making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2), 36-46.
- Gersh, J., Lewis, B., Montemayor, J., Piatko, C., & Turner, R. (2006). Supporting insight-based information exploration in intelligence analysis. *Communications of ACM*, 49(4), 63-68.
- Harper, B.D., & Norman, K.L. (1993). Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5, *Proceedings of Mid Atlantic Human Factors Conference*.
- Hsieh, H., & Shipman, F.M. (2002). Manipulating structured information in a visual workspace, 15th annual ACM symposium on User interface software and technology (pp. 217-226). Paris, France: ACM Press.
- Huang, X., & Soergel, D. (2006). An evidence perspective on topical relevance types and its implications for exploratory and task-based retrieval. *Information Research*, 12(1), 281.
- Lowrance, J.D., Harrison, I.W., & Rodriguez, A.C. (2001). *Capturing Analytic Thought*, *Proceedings of the 1st International Conference on Knowledge Capture*. Victoria, British Columbia, Canada: ACM Press.
- Pirolli, P., & Card, S.K. (2005). The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis, *International Conference on Intelligence Analysis*.
- Qu, Y., & Furnas, G.W. (2007). Model-Driven Formative Evaluation of Exploratory Search: A Study under a Sensemaking Framework. In R.W. White, G. Marchionini & G. Muresan (Eds.), *Evaluating Exporatory Search Systems: Special Issue of Information Processing and Management* (Vol. To appear).

- Savolainen, R. (2006). Information use as gap-bridging: The viewpoint of sense-making methodology. *Journal of the American Society for Information Science and Technology*, 57(8), 1116.
- Stefik, M.J., Baldonado, M.Q.W., Bobrow, D., Card, S., Everett, J., Lavendel, G., et al. (1999). The knowledge sharing challenge: The sensemaking white paper: PARC, Inc.
- Vakkari, P. (2000). Relevance and Contributing Information Types of Searched Documents in Task Performance, *SIGIR*: ACM Press.
- Wang, W., & Haake, J. (1997). Supporting user-defined activity spaces, *Proceedings of the eighth ACM conference on Hypertext* (pp. 112-123). Southampton, United Kingdom: ACM Press.
- White, M.D. (1975). The communications behavior of academic economists in research phases. *Library Quarterly*, 45, 337-356.
- Wright, W., Schroh, D., Proulx, P., Skaburskis, A., & Cort, B. (2006). In The Sandbox for Analysis - Concepts and Methods (pp. 801-810). Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada. ACM Press.

Support Sense-Making with Tools for Structuring a Conceptual Space

1. Introduction

This abstract describes the design of a sense-making tool that assists users to structure their conceptual space and create a task report from the conceptual structure. Sense-making involves the

- recognition of a knowledge gap,
- seeking for information,
- analyzing and synthesizing information to create an understanding, and possibly
- producing a task output: a report, decision, or other type of output.

A typical sense-making task is for an intelligence analyst to gather, analyze, and synthesize information related to some political figure or event, and to make recommendations for action. Everyday sense-making tasks, although less complex, share certain characteristics with the example.

Many people use information systems as sense-making tools. Standard information retrieval systems can support reasonably well the search for pieces of relevant information when the user can identify her information need or knowledge gap to certain extent. However, in order to make sense of the information they found, users need to understand the pieces of interrelated information, identify patterns, and build on their previous knowledge to create an updated understanding.

Most sense-making tasks are characterized by the interplay of searching for information and creating a structured representation of the situation, problem, or domain. In difficult sense-making tasks, sense makers often use external more or less structured representations to store the information for repeated manipulation and visualization (Stefik et al., 1999). Users need sense-making tools that facilitate the creation of such conceptual structures.

Research in the areas of education (especially learning theories), cognitive psychology, and information extraction all bring useful insights into sense-making research. This study aims to integrate findings from these areas and build on that basis to design a sense-making tool that assists users in creating and using external representations for sense-making. Some questions to be explored are:

1. How do users build the concepts and relationships in their conceptual models?
2. How can the tool support the co-existence and transformation of different formats in which users represent their mental models of concepts and relationships?
3. How can automatic or computer-assisted extraction of entities and relationships assist users in building a representation?
4. Does the framework of topical relevance types assist users in making connections between retrieved information and conclusions to be reached and in identifying types of information to search for?

2. Related Research

2.1 Sense-making

Sense-making is characterized as a series of continuing gap-bridging activities (Dervin, 1998; Savolainen, 2006). Through cognitive task analysis, (Pirolli & Card, 2005) proposed a notional model of sense-making, with two loops of activities:

- (1) a foraging loop that involves seeking and extracting information, and
- (2) a sense-making loop that involves iterative development of a conceptualization.

For the information foraging loop, researchers identified the important role of exploratory search and developed systems to support it (Qu & Furnas, 2007). The sense-making loop, on the other hand, including activities such as skimming, examining details, summarizing, organizing, and identifying patterns, is not as well supported. A key task in sense-making is to identify patterns of concepts and relationships to build on. Structured representations play an important role in this task.

Tools have been developed to support sense-making in various ways, mostly to capture intermediate products of sense-making such as insights (Gersh, Lewis, Montemayor, Piatko, & Turner, 2006) and analytical thoughts (Lowrance, Harrison, & Rodriguez, 2001), and to provide a workspace of the intermediate representations. (Hsieh & Shipman, 2002; Wang & Haake, 1997; Wright, Schroh, Proulx, Skaburskis, & Cort, 2006) However, there is less support for connecting intermediate products to the conceptual structure that users develop. This research aims to develop a workspace that supports users' structuring a conceptual space using and consisting of various sources, including search results and intermediate structured representations such as concept maps, templates, and outlines.

2.2 Information Extraction

Automated extraction of concepts and relationships helps users with pieces of useful information from various sources. It may suggest preliminary formal statements for users to examine and filter, saving users the time reading the documents and extracting relationships manually. A flexible sense-making tool should support the co-existence and transformation (Wang & Haake, 1997) of information structures in different degrees of formality. How to organize and integrate extracted results into the emerging conceptual structure of users remains a question to be investigated.

2.3 Task-based Information Seeking and Relevance Research

Sense-making is often embedded in tasks. Task-based information seeking research (Vakkari, 2000; White, 1975) found that different types of information are sought at different task stages, for example, background information are sought at the beginning (e.g. pre-focus) stage of the task. Research in relevance (Huang & Soergel, 2006) reveals different ways in which a piece of information may be useful to a task. This paper aims to build on the finding to investigate how the different relationships between a piece of information and a task may be used in a sense-making tool to help users throughout the different stages of sense-making, especially how to organize

different sources and formats of information based on these relationships in the visual workspace for creating a conceptual structure.

3. Design of a Sense-making Tool

This section describes the design of a sense-making tool to assist users in building their conceptual model of a task situation by organizing their search results, identifying and recording concepts and relationships, and outlining a task report for further use. It is to be integrated with a standard IR system. We envision the following major functions:

1. Searching

- Users can use the regular search box to issue a query or they can initiate a search from the workspace on a particular part of the conceptual structure.
- Users can search for entities and relationships extracted by the system with different levels of manual involvement (shown in Figure 1).

Search		Workspace	
<input type="text" value="Al-Bashir"/>		Type in your query or right click on a concept to start a search	
Source text		Extracted Relationships related to Al-Bashir	
President of the Republic, Field Marshal Omar Al-Bashir has affirmed that Sudan is considered.. <i>Suna News Agency – July 20, 2006</i>		<i><is president of> Sudan</i>	
... Back in Sudan Al-Bashir led a series of successful assaults on the SPLA in ... <i>BBC news - Oct 1, 2001</i>		<i><fights against> Sudan People's Liberian Army (SPLA)</i>	
President Al-Bashir has lauded the efforts of sister Chad for boosting the security and stability in Darfur states... <i>BBC news - Sep 11, 2003</i>		<i><is supported by> Chad</i>	
Egyptian Prime Minister held talks in Cairo late Sunday ... At the beginning of the Egyptian-Sudanese higher committee meeting ... <i>BBC news - Jul 28, 2002</i>		<i><met with> Egyptian Prime Minister</i>	

Figure 1: Search results in formats of source text and extracted relationships

2. Recording and manipulation of interim representations of concepts and relationships in concept maps, templates, and outlines:

- Users can create, modify, and delete concepts or relationships based on useful search results or previous knowledge, shown in Figure 2.
- Users can attach a piece of evidence found in any text segment and its citation information to the relevant parts of their conceptual model, and specify the relevance type.

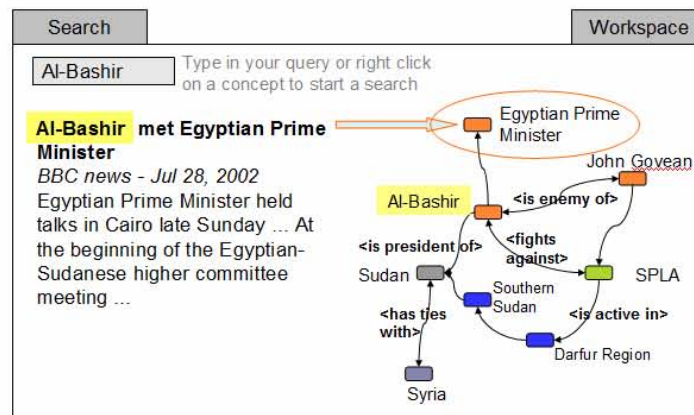


Figure 2: Creating an entity from search results into an existing concept map

3. Supporting co-existence and transformation of structures in different degrees of formality: users can switch between graphic, template-based, and outline displays of concepts and relationships.
4. Organization of search results and creation of task reports: users can annotate useful search results and categorize them by subject or task component and connect them according to the relevance types identified by (Huang & Soergel, 2006). Such categories may be quite useful to users throughout the sense-making. At the beginning stage, background information is needed. When creating a task report, comparison of similar situations and how they were handled may be especially useful. The relevance types may provide guidance in what to search for next (for example, information on similar situations for comparison). The display and organization of search results and of the information in the work space can also differentiate what is direct evidence to an argument and what are inferred by the user based on indirect evidence based on user input from an earlier stage.

4. Iterative Design and Evaluation

The research design involves iterative design and evaluation of the sense-making tool for structuring a concept space. Participants will be 16-20 journalism and political science students. They will work on two assigned tasks and one task of their own. Each task session takes about 90 minutes, 30 minutes for a brief introduction of the tool and a training task, 60 minutes for the main task.

Data collection will involve:

- Two pre-session questionnaires/interviews to learn about user background and background knowledge about the task;
- A post-session Questionnaire for User Interaction Satisfaction (QUIS) (Harper & Norman, 1993) (modified) to learn about how users think about the tool;

POSTER SUBMISSION 6:

Older adults and information technology: The current state of research and future directions

AUTHOR(S):

Birkland, Johanna L.H.

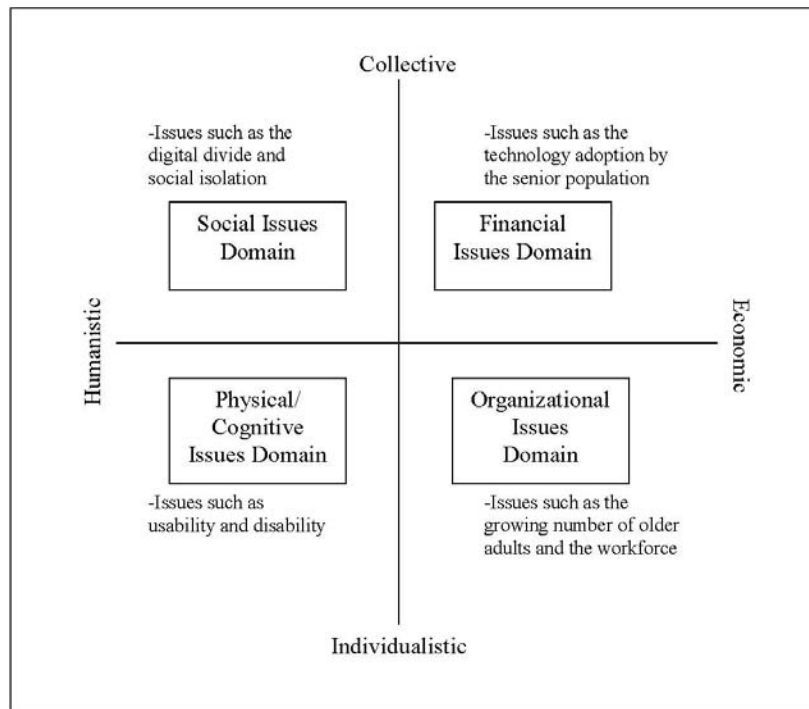
Kaarst-Brown, Michelle L.

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Figure 1. Domains of issues addressed in the literature regarding older adults and information technology



References

- Hedge, J. W., Borman, W. C., & Lammlein, S. E. (2006). *The aging workforce: Realities, myths, and implications for organizations*. Washington, D.C.: American Psychological Association (APA).
- U.S. Census Bureau. (2005). Facts for Features: Older Americans Month Celebrated in May. Retrieved December 3, 2006, from http://www.census.gov/Press-Release/www/releases/archives/facts_for_features_special_editions/004210.html

Older adults and information technology: The current state of research and future directions

Poster Submission

In the US alone, it has been predicted that the senior population will grow to represent over 21% of the total population within the next fifty years, a 147% increase in the number of seniors from 2000 (U.S. Census Bureau, 2005). With the aging of the population and the seriousness of issues that face older adults, it will be important to discover ways in which information technology can be used to benefit older adults' lives and society as a whole. Research concerning older adults and information technology is dispersed across several disciplines and a variety of perspectives, with the majority of information systems and information science literature focusing on only a few of the issues that older adults and society in general will face in the upcoming decades.

This poster will describe the authors' work in synthesizing research regarding older adults and information technology across several different disciplines into a congruent conceptual framework. The purpose of this work is to help researchers to understand the breadth and depth of the issues that have been researched in regards to this area. In addition, the authors' propose a research agenda for the fields contained within iSchools regarding older adults and information technology.

Based upon a review of the literature, we have conceptualized the issues that have been explored into four major domains of current research: social concerns, financial concerns, physical and cognitive concerns, and organizational concerns facing older users and the use of information technology. Examples of social concerns regarding usage of

technology include issues such as the digital divide and the use of computers to lessen social isolation. Financial concerns have included studies on technology adoption and marketing to older individuals. Studies regarding the growing percentage of older adults in the population and the impacts that this will have upon organizational knowledge retention and technology retraining represent issues in the organizational concerns domain. Physical and cognitive concerns have examined issues such as disability and assistive technology. The authors suggest that the social and physical/ cognitive issues remain the most addressed domains in the study of older adults and technology.

As can be seen in the diagram below, social and financial concerns have been addressed from a collective perspective, which has typically portrayed these issues as systematically impacting the entire older adult population and also having broad repercussions that affect society as a whole. For instance, the digital divide, a social concern, is often portrayed as systematically affecting all older adults and society. Organizational and physical/ cognitive concerns have been addressed from an individualistic standpoint and have typically been portrayed as impacting our society from the individual level. For instance, issues such as dependency of older adults have been explored through the individual level, where solutions have been proposed that facilitate the dependency needs of single individuals. This relationship between concerns that have been addressed from an individual or collective standpoint is represented by the vertical axis of the figure.

Beyond conceptualizing these issues along a collective and individualistic continuum, the various issues can also be conceptualized along a continuum of economic and humanistic issues. The horizontal axis represents this relationship. As can be seen,

social and physical/ cognitive concerns have typically been addressed from a humanistic standpoint, where it is suggested that these issues represent societal impacts due to decreased older adult participation in our society. For instance, social isolation of older adults is often viewed as negatively impacting not only the seniors who suffer such isolation, but also society as these adults' contributions are removed through isolation. Issues that have been considered from a more humanistic standpoint have often considered the problems that older adult's face through an activist lens, which sees technology as a potential tool for empowerment of older individuals. Issues that have traditionally been explored from an economic perspective involve concern over potential loss of revenue for organizations, revenue retention as the population ages, or developing sources of revenue from the older population. For instance, issues such as technological adoption by seniors have typically examined older adults as a marketable population. Issues that have been considered from a more economic standpoint have often considered the problems that older adult's face as being economically created or consider the issues surrounding the aging of our society as an economic problem.

Very few issues regarding older adults and technology have been addressed by researchers from the fields traditionally contained within iSchools. Information systems research in this area has mainly focused on exploring issues in the cognitive and physical domain, with most research concentrating on usability issues surrounding information technology use by the elderly. Information science research in this area has focused on several of the social concerns, including the digital divide. The authors suggest several directions for potential iSchool contributions to this area of research, including a

diversification of the methods used to study each domain of issues and exploring the impact of emerging forms of technology on older individuals.

As our population continues to age, our entire society will be faced with the issues of equal access, disability, healthcare needs, engagement, and technological literacy of older adults. Our organizations will be pressed to determine ways to manage retiree's knowledge and with ways to accommodate older workers who choose to work beyond our age of retirement (Hedge, Borman, & Lammlein, 2006). Additionally, as our society becomes more technologically centered, it will be important to explore ways in which we can improve and maintain the technological literacy and access of older adults so that they may remain active in our political and social systems. The authors suggest that the exploration of many of these issues would benefit from the iSchools' unique interdisciplinary lens.

POSTER SUBMISSION 7:

Why leave Wikipedia?

AUTHOR(S):
Jian, Lian

MacKie-Mason, Jeffrey K.

ORGANIZATION(S):

University of Michigan, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work, pages 1–10, New York, NY, USA, 2005. ACM. ISBN 1-59593-223-2. doi: <http://doi.acm.org/10.1145/1099203.1099205>.

Yan Chen, Max Harper, Joseph Konstan, and Sherry Xin Li. Social comparisons and contributions to online communities: A field experiment on movielens, 2007. Working paper.

Daniel Hughes, Geoff Coulson, and James Walkerdine. Free riding on gnutella revisited: the bell tolls? *Distributed Systems Online*, 6(6), June 2005. doi: 10.1109/MDSO.2005.31.

Aniket Kittur, Ed Chi, Bryan A. Pendleton, Bongwon Suh, and Todd Mytkowicz. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007)*, April 28 - May 3 2007.

Andrew Lih, 2007. URL <http://www.andrewlih.com/blog/2007/06/28/wikipedia-plateau/>.

Oded Nov. What motivates wikipedians, 2007. URL <http://opensource.mit.edu/papers/Nov.Wikipedia.motivations.opensource.mit.edu.pdf>. Accepted for publication in the Communications of the ACM.

Reid Friedhorsky, Jilin Chen, Shyong (Tony) K. Lam, Katherine Panciera, Loren Terveen, and John Riedl. Creating, destroying, and restoring value in wikipedia. In *GROUP '07: Proceedings of the 2007 international ACM SIGGROUP conference on Supporting group work*, New York, NY, USA, 2007. ACM.

Richard Ryan and Edward L. Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68–78, 2000.

Why leave Wikipedia?

Poster Submission

October 29, 2007

1 Introduction

Some user-contributed content (UCC) applications, such as Yahoo!Answers, Wikipedia, and YouTube have drawn much media attention. Various reasons motivate the tremendous contributions to a few UCC systems so far. Existing literature has uncovered many factors that affect users' decisions to *become* contributors [Bryant et al., 2005], to *continue* contributing [Nov, 2007], and to *increase* contribution [Chen et al., 2007], but none of them pay attention to why active contributors decide to *stop* contributing. The exit of active contributors may affect quantity and quality of content provision on UCC systems.

Indeed, preliminary evidence shows that there is a reason to worry about the long term sustainability of some systems. In 2000, Adar and Huberman's study on the Gnutella network showed that there was high level of free-riding on this network. Five years later, Hughes et al. [2005] find on the same network free-riding has gotten worse: the percentage of users who do not share any files has increased from 66 % to 85%. In the mean time, Andrew Lih, one active Wikipedia research/contributor also blogged about its growth rate slowing down dramatically in late 2006 [Lih, 2007]. To what extent is this due to contributors leaving the system? If so, who left? And for what reasons? Answers to these questions are useful to UCC system designers for determining the impact of contributors leaving or for devising mechanisms to prevent them from leaving.

In this study we focus on one system, Wikipedia, and analyze why some editors stop contributing. We chose Wikipedia for a number of reasons. First, it has a large number of active contributors. By mid-2006 over 10,000 editors had made more than 100 edits [Kittur et al., 2007]. Second, Wikipedia maintains a detailed record of its contributors' activities, and shares its database online.¹

2 Method

We will conduct statistical analysis on a dataset we obtained from the English Wikipedia on July 16, 2007. This dataset contains the full edit history of every article and user page on Wikipedia. We take a simple random sample of all the registered editors. In

¹See http://en.wikipedia.org/wiki/Wikipedia:Database_download.

Table 1: Predictors of Wikipedians' Decisions to Exit.

	Name
Independent variables	creator dummy
	preserver dummy
	destroyer dummy
	<i>D_OnTime</i>
	<i>D_Deled</i>
	work intensity
	article stability
Controls	total lifetime number of edits

our study, an editor is said to "leave Wikipedia" if they do not edit for three months or longer. The choice of three months as a cut-off is arbitrary. In the following sections we explain how we operationalize other variables, and state our hypotheses.

2.1 Lifespan of a Wikipedian

First, we ask how the lifespan of a Wikipedian is distributed. We estimate this distribution using a survival analysis with truncated data — treating the active Wikipedians' data points as truncated² — while assuming a parametric distribution function such as lognormal or Weibull. This distribution provides a baseline for exit behavior. We can use this distribution to estimate, for example, the effects of changes in exit timing.

2.2 What predicts a Wikipedian's departure?

Next, we identify factors that predict a Wikipedian's departure and estimate their effects. We analyzed four hypothesized predictors: roles played in the system, peer feedback, work intensity, and article stability. The effects of these predictors will be estimated simultaneously in a logit regression, controlling for the editor's total number of edits. Table 1 lists our independent and control variables.

For each observed editor, we define three "decision periods": *lifetime*, *last week*, and *last month*. The lifetime of an editor begins with her first edit and ends with her last edit in our dataset. If an editor is deemed as having "left" Wikipedia, her *last week (month)* is the last week (month) of her participation, not the last week (month) in our dataset. For a currently active editor, the *last week (month)* refers to the last week (month) in our dataset.

2.2.1 Roles

We identify three different roles among Wikipedians: creator, preserver, and destroyer. An editor plays one or more of these roles depending on the type of edits she contributed.

²That is, their observed lifespans are shorter than their actual lifespans due to the cutoff date of observation.

We classify all edits into four mutually exclusive types: creation, reversion, deletion, and damage. Reversions are easy to identify: if the texts are exactly the same as an earlier version, judged by the MD5 checksums of the texts, it is a reversion. With the help of techniques introduced by Friedhorsky et al. [2007], we can identify damages. First, edits that are reverted in future are identified as will-be-reverted (WBR) edits. Damages are WBR edits “where the future reverts edit comment suggests either (a) explicit intent to repair vandalism or (b) use of revert-helper tools or autonomous anti-vandalism bot activity.” If an edit is not classified as a reversion or a damage, it is classified either as a creation or as a deletion. If an edit increases the total number of words in the text or it adds new words into the text, it is a creation. On the other hand, if an edit decreases the total number of words in the text and no new words are added, it is a deletion. Suppose editor i has edited w_i words in her lifetime, within which w_{ci} were created, w_{ri} were involved in her reversions, w_{di} were deleted, and w_{mi} were damages. We calculate the proportion of her creation as follows,

$$p_{ci} = \frac{w_{ci}}{w_{ci} + w_{ri} + w_{di} + w_{mi}}; \quad (1)$$

and her proportion of reversions and deletions as follows,

$$p_{rdi} = \frac{w_{ri} + w_{di}}{w_{ci} + w_{ri} + w_{di} + w_{mi}}; \quad (2)$$

and lastly, her proportion of damages as follows,

$$p_{mi} = \frac{w_{mi}}{w_{ci} + w_{ri} + w_{di} + w_{mi}}. \quad (3)$$

Let p_c denote the population mean of p_{ci} , p_{rd} the population mean of p_{rdi} , and p_m the population mean of p_{mi} . If $p_{ci} > p_c$, editor i is labeled as a creator. If $p_{rdi} > p_{rd}$, editor i is labeled as a preserver. And if $p_{mi} > p_m$, editor i is labeled as a destroyer. We do not have strong prior hypotheses regarding how editor roles affect departure. Our findings will inform us about the effect of exit behavior on some aspects of the quality of UCC contributions.

2.2.2 Recent peer feedback

It is rewarding to see ones' own edits persist. It may also be disappointing to see ones' own edits being removed by others. How other editors react to one's edits serve as feedback, and people seek self-satisfaction from peer feedback [Ryan and Deci, 2000]. In addition, Wikipedia makes it easy to watch peers' feedback by providing a watchlist function, so that a registered editor can monitor changes made on any page, certainly including the pages that she edits.

We define two variables, *Ontime*—the number of minutes that a Wikipedian's edits persist, and *Deled*—the number of times a Wikipedian's edits get deleted. We hypothesize that if *Ontime* increases for an editor, she is rewarded and is motivated to contribute more, hence is less likely to quit. On the other hand, if *Deled* increases, she is more likely to leave. For editor i , we denote the changes in these variables by D_Ontime_i and D_Deled_i , and derive them as follows,

$$D_Overtime_i = Overtime_i(\text{last week}) - Overtime_i(\text{week before last week}) \quad (4)$$

$$D_Deled_i = Deled_i(\text{last week}) - Deled_i(\text{week before last week}) \quad (5)$$

Our hypotheses are that the probability of leaving decreases in $D_Overtime_i$, and increases in D_Deled_i .

2.2.3 Last month work intensity

Within the Wikipedia community some speculate that getting burnt-out causes some editors to quit [Lih, 2007]. Thus those editors who have left may have worked harder than usual right before leaving. We propose to use the following formula to describe the work intensity in the last month of editor i ,

$$\text{Last month work intensity}_i = \frac{\text{number of edits in the last month}}{\text{lifetime average number of edits per month}} \quad (6)$$

We hypothesize that the higher a Wikipedian's work intensity, the more likely that she will leave.

2.2.4 Last week article stability

Some Wikipedians are only interested in editing a particular set of pages. When these pages reach a stable stage, they may find little motivation to continue. We identify the articles an editor cares about by looking at how much she has contributed to them. Suppose editor i has contributed to n articles in total, with e number of edits to each of these articles on average. We define all articles which have received more than e edits from editor i as the set that i cares about. The stability of an article is measured by the number of edits it receives in the last week of editor i . The stability of the set of articles is the average stability of each article in the set. We hypothesize that the more stable the articles that an editor cares about, the more likely that this editor will stop contributing.

Acknowledgments

This section is blanked out for blind peer review purposes.

References

- Eytan Adar and Bernardo A. Huberman. Free riding on Gnutella. *First Monday*, 5, 2000.
- Susan L. Bryant, Andrea Forte, and Amy Bruckman. Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *GROUP '05*:

POSTER SUBMISSION 8:

Interactive Machine Learning (IML) Markup of OCR Generated Text by Exploiting Domain Knowledge: A Biodiversity Case Study

AUTHOR(S):
Wei, Qin

Heidorn, P. Bryan

ORGANIZATION(S):
UIUC, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Reference:

David Robins. Interactive Information Retrieval: context and basic notions. *Information Science (Special issue on information science research)*, 3(2):57-61, 2002.

Hong Cui. Automating semantic markup of semi-structured text via an induced knowledge base: a case-study using floras. Ph.D Dissertation. University of Illinois at Urbana-Champaign. 2005.

James R. Curran. Blueprint for a high performance NLP Infrastructure. In *Proceedings of the HLT-NAACL 2003 workshop on Software Engineering and Architecture of Language Technology Systems*. 8:39-44, 2003.

Jurgen Koenenman, Nicholas Belkin. A case for interaction: a study of interactive information retrieval behavior and effectiveness. In *Proceedings of CHI'1996*, 205-212, 1996.

L. Venkata Subramaniam. et al. Information extraction from biomedical literature: methodology, evaluation and an application. In *the Proceedings of the Twelfth International Conference on Information and Knowledge Management*. New Orleans, L.A. Pages:410-417, 2003.

Malcolm Ware, et al. Interactive machine learning: letting users build classifiers. *International Journal of Human-Computer Studies*. 56(3): 281-292, 2002.

Reed S Beaman, et al., HERBIS: Integrating digital imaging and label data capture for herbaria. *Botany 2006*, Chico, CA. July 28-August 2, 2006.

Rupesh R. Mehta. et al. Extracting semantic structure of web documents using content and visual information. *Special interest tracks and posters of the 14th international conference on World Wide Web WWW '05*. Chiba, Japan. Pages: 928-929, 2005.

Special issue on interactivity at the Text REtrieval Conference (TREC). *Information Processing and Management*, 37(3), 2001.

Tom M. Mitchell. Machine learning. McGraw Hill, 1997.

Vinay Borkar, Kaustubh Deshmukh, Sunita Sarawagi. Automatic segmentation of text into structured records. *ACM SIGMOD*. 30(2): 175-186, 2001.

Interactive Machine Learning (IML) Markup of OCR Generated Text by Exploiting Domain Knowledge: A Biodiversity Case Study

Several digitization projects such as Google books are involved in scanning millions of books. The Biodiversity Heritage Digital Library (BHL <http://www.bhl.si.edu/>) plans to scan 1 million volumes of biodiversity literature over the next five years. However, the usefulness of the scanned images is limited because they can only be accessed through existing catalog information. Images can not be easily manipulated and transformed to useful information in full-text information systems. "Because of the very large amounts of data being generated, it is difficult to have human curators extract all these information and present them in a form useful to researchers. Information Extraction (IE) from such sources is becoming crucial for the timely dissemination of information." (Subramaniam, 2003). Consequently, simple approaches that transform the text to structured format such as XML or relational databases will not be successful.

Machine Learning (ML) techniques, especially Supervised ML (SML) have been used widely in information extraction (IE) and automatic markup. "ML has proven to be of great practical value... They are especially useful in (a) data mining problems where large databases may contain valuable implicit regularities that can be discovered automatically..." (Mitchell, 1997). IE and automatic markup of the biodiversity documents is this kind of domain. Substantial research has been conducted on the usefulness of ML in IE and automatic markup (e.g. Borker 2005; Cui 2005). Borker demonstrated 87% F-score in automatically extract address elements (eg. house number, street name, city) from addresses and bibliographic entries from bibliography resources. Cui's dissertation (2005) demonstrated that domain knowledge gained from machine learning models in one publication is very useful for improving the performance of automatic markup in another publication in the same field.

One of the least tapped sources of biodiversity knowledge is the collection locations, dates, species identification and other information on over a billion natural history specimen labels worldwide. Only a very small fraction of these have been digitized and the information added to databases (Beaman et al., 2006). The HERBIS (<http://www.herbis.org>) project has build tools to allow researchers to submit images of these specimens to a web service and receive an extended Darwin core document¹ in return. Using the Herbis Learning System (HLS), we extract 36 independent elements of information from these labels. The automated text extraction tools are provided as a web service so that users can reference digital images of specimens and receive back an extended Darwin Core XML representation of the content of the label. The classification of the sub-elements is accomplished using SML. A training set was constructed using a collection of 145 examples which contains 4183 element classifications. The dataset comes from digitized OCR records from the Yale Peabody Herbarium with multiple label formats randomly selected from the type written labels and OCRed by ABBYY software. We coded the data as a Relax NG Schema allowing all elements occurrence to be

¹ See <http://wiki.tdwg.org/wiki/bin/view/DarwinCore/DarwinCoreDraftStandard>

optional, potentially occurring multiple times and in any order as is required by the variability in the input data. The relaxNG schema could be found online².

Many text classification ML algorithms are available such as: Naive Bayes, Hidden Markov Model, Decision Trees, Support Vector Machines. Each algorithm has its own advantages and shortcomings. The properties of our data helped us select particular algorithms. Museum labels have a relatively loose sequential structure, a high level of OCR errors, some fields have restricted sets of possible fillers while others are “open world” and may contain almost any text. For our tasks, a few of the fields are more important than others such as: family, genus, species, collector, and date. Several experiments need to be carried out to test several promising candidate algorithms and analysis their potential benefits and limitations of using them. For static evaluation, f-scores and ten-fold validation can be used. Because of the structure of the data we implemented a modified Hidden Markov Model and Naïve Bayes Model. A Hidden Markov Model (HMM) consists of states (in our case, they are the different kinds of elements), observations, start probability, transition probability and emission probability. Each state emits one or more symbols in the dictionary from a probability distribution for that state. Beginning from the start state, a HMM generates an output sequence by making transitions from one state to the next up to the end state. So the HMM model is an order preserving algorithm which is the primary feature of this model. It is currently widely used in web-content mining and speech recognition. A Naïve Bayes (NB) model is a probability model based on conditional probabilities. NB model make predictions based on the probability distribution of features from the training set. NB then uses the distribution information to calculate the probabilities of a new instance belong to the classes. The example would then be classified to the highest probability class. NB model has been proven good in some problems both in data mining and text mining. The performance of both models in our dataset could be found in figure 1.

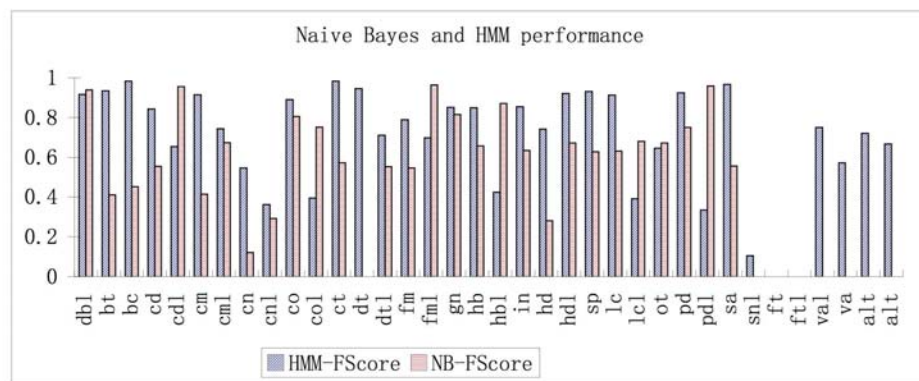


Fig. 1 Performance of NB and HMM

² <http://www3.isrl.uiuc.edu/~TeleNature/Herbis/semantirelax.rng>

The results from the two unintegrated algorithms are encouraging. Generally, NB performs better than HMM on elements that are “labels” or markers for other labels. All these codes end with a letter “I”. For example taxonomic family is coded as “fm” and family labels are “fml.” Performance could be improved by integrating the algorithms, using the best algorithm for individual labels. MorphBank³ and some other projects are coordinating with the HERBIS development teams to provide an automatic markup module for museum specimen digitization projects. We are expanding our research to use a more active architecture, Interactive Machine Learning (IML) as introduced by Ware, et al. 2002. Currently most ML systems are built by computer scientists (programmers) using expert generated data, not the domain experts. In the standard (non interactive) ML procedure, building a learner/classifier is a fully automated process. As Cui demonstrated most ML systems do not fully take advantage of the domain knowledge which could be very beneficial if used properly. IML “offers a natural way of integrating background knowledge into the modeling state.” (Ware, 2002).

Future Work:

System Design and Implementation: Unlike traditional ML, IML is a “human-in-the-loop” system. The system would be initialized with one of previously constructed models for one or more ML algorithms such as NB and HMM. A person using the system for the first time would feed raw museum label images through these models, which in turn would return the labels marked up in XML. Using a graphical user interface which represents the XML with more user friendly color coding, the user corrects any errors in the machine classification. The system can use these new label instances as a new training set to create new ML models tailored specifically to this users data. Given a sufficient number of examples, the performance of these new models should exceed the performance of the generic models that came with the system. With each batch of new records that the user submits, the system gives the user feedback on the relative performance of the available ML models. As the system performance increases the number of corrections that the user needs to make decreases. In essence each user can tailor a personalized IML system. By sharing the resulting models with other users, we have a social network of IML modules. The machine learning components of the system will be provided as a web service so that other people can build other interfaces over the IML web service modules.

³ <http://www.morphbank.net/>

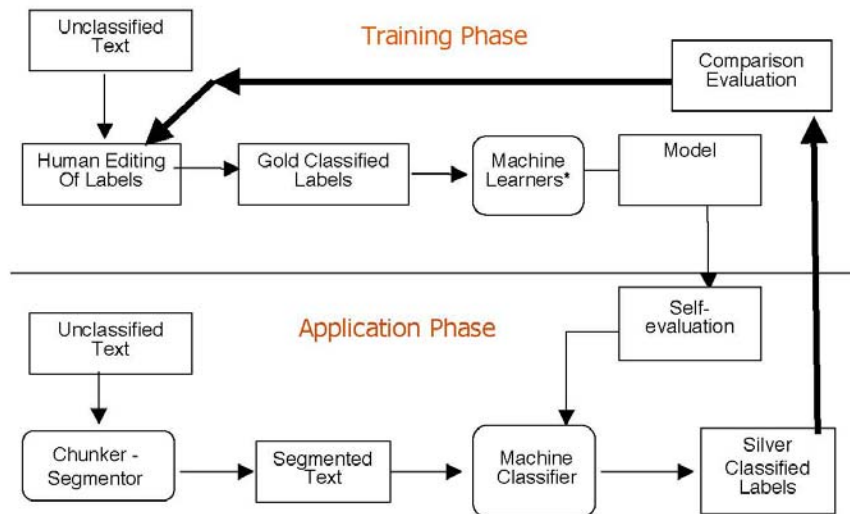


Fig 2. Interactive Machine Learning Architecture (*Machine Learners" in the diagram should be a stack of overlapping learners depending on which one the user selected.)

User centered experiments and data analysis. Standard precision, recall and F-Scores are not sufficient for evaluating interactive systems. While IML is fairly new, both Interactive Information Retrieval (IIR) and Interactive Data Mining (IDM) have been studied extensively. The importance of IIR could be found in TREC tasks. It appeared since the first TREC interactive query mode (TREC-1,2), interactive track (TREC 3-8), Manual query mode (TREC 1-7), and high precision (TREC-6,7). Our vision of the user-end experiments would be similar as the experiments done in IIR in TREC. The focus would be studying user's behavioral details, the process, and interim results as well as the summary of final results and the effects of the system, searcher and their interactions. Important variables are the number of human corrections required per some number of records, the time required to correctly complete a fixed number of labels, number of training examples and number of error corrections needed to meet some performance criteria such as a 90% F score. But we would investigate several more measures that would be more suitable for Machine Learning and Automatic Markup. We will identify and discuss why we chose the measure and what's the advantages and limitation of each measure.

POSTER SUBMISSION 9:

Proposal of Document Classification with Word Sense Disambiguation

AUTHOR(S):
Liu, Xiaozhong

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

1. Background

An important area of NLP is the study of Word Sense Disambiguation (WSD), which may assign a unique word sense to a word. There are different methods to implement WSD: one is the word sense based on the collocation of other words (Yarowsky, 1993), where nearby words provide strong consistent clues to the sense of a target word, conditional on relative distance, order and syntactic relationship; and the other is the word sense based on discourse (Gale et al, 1992), where the sense is consistent within any given document. Many experiments in recent years of both supervised (Leacock 1993) and unsupervised (Yarowsky, 1993) WSD algorithms have accomplished promising performance with a high precision rate.

Another important area in the field of text mining (Lewis & Spark Jones papers) is document classification, which identifies one or more of several topic labels for a text document. A significant body of research has improved the results of document classification, with innovations in identifying document features as well as improving algorithms.

In this study, we will use WSD as part of a method to create innovative features to represent the documents for classification task. With the help of WSD, a set of specially selected ambiguous words can be further distinguished by word sense clustering, in order to achieve better document classification.

2. Hypothesis

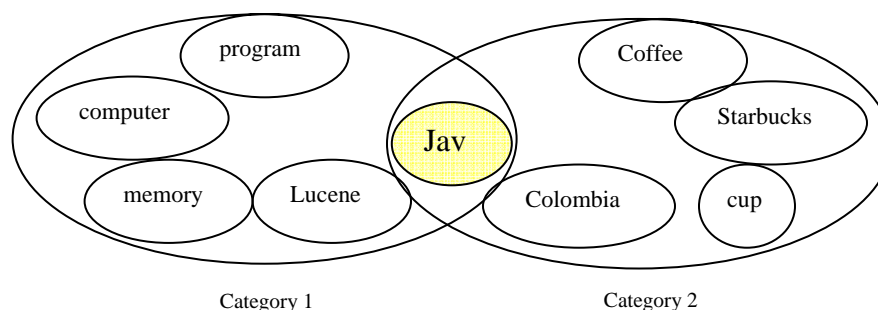
The existing methods for document classification (Lewis 1992) are mainly focused on bag of words and phrases as features to train a classification system with machine learning algorithms such as Support Vector Machines (SVM). More recent work, such as that in (Yilmazel 2006) adds semantic features, such as Part-Of-Speech tags, to improve the classification.

However, using individual words as features could lead to ambiguity in the features; this is the polysemy effect (Scott 1990). As an example of this effect, supposed that Document A includes the words “Java”, “memory” and “computer”, while Document B has words “Java”, “Coffee” and “Starbucks”. If we use words as features to classify the document, it might mislead the system by using the word “Java” to classify the two documents into the same group. However, if we can target “Java” as an ambiguous word, we could use collocation level or discourse level WSD to disambiguate this word into multiple features, one for each of the different senses of the word in the collection.

Three hypotheses are listed here:

H1: Considering the success of WSD to distinguish word senses which can solve the polysemy dilemma, there will be chance to improve the performance of document classification by using word senses as feature.

H2: It is unnecessary to disambiguate every word in the document to use its senses as features, because firstly WSD is expensive and not very accurate; secondly too much word sense disambiguation may increase the system's risk of “overfitting”. So the better way is to select a subset of words which are the most informative and distinctive in the collection of ambiguous words. The training corpus for document classification already contains class label information, and a word which frequently appears in different classes could confuse the classification algorithm and should be disambiguated. Words, such as “Java” in Fig 1, shared by different categories of class labels (not stop words) are likely to be the most ambiguous words.



identify the most ambiguous words between categories of class labels

H3: Lexical resources such as WordNet, have very subtle distinction between word senses, which could be harmful for document classification. (R Mihalcea 2001 & WB Dolan 1994) For example, WordNet distinguishes between bass -- (the lowest part of the musical range) and bass or bass part -- (the lowest part in polyphonic music). We should cluster the similar word senses together as the same feature before input into the classification algorithm.

3. Methodology

Several innovative methods are proposed in this paper to utilize word senses in the features for document classification.

3.1 Identify confusion set of words

As discussed above, if we disambiguate all the possible words in the corpus, we will take the risk of overfitting and negatively affect the classification performance. So, the first step is to identify a subset of words in the corpus that need to be disambiguated, namely, the subset of words that most likely confuse the classification algorithm. We call this the confusion set.

(Yang et al 1997) mentioned that information gain is one of the most effective methods for the classification task. If we have m different categories (from C_1 to C_m), the information gain of a specific term t could be defined as:

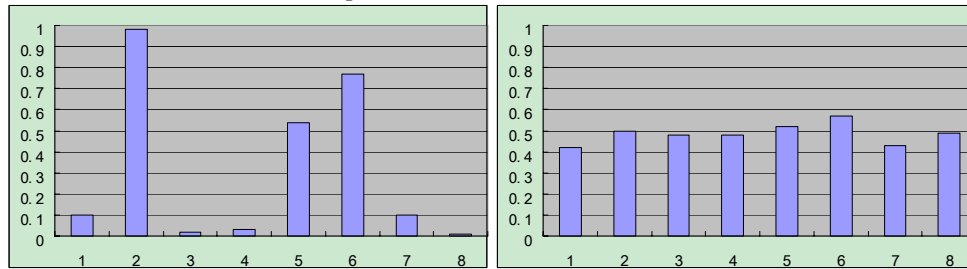
$$IG(t) = H(C) - H(C|t) - H(C|\bar{t})$$

$$= -\sum_{i=1}^m P(C_i) \cdot \log P(C_i) + P(t) \sum_{i=1}^m P(C_i|t) \cdot \log P(C_i|t) + P(\bar{t}) \sum_{i=1}^m P(C_i|\bar{t}) \cdot \log P(C_i|\bar{t})$$

$P(C_i)$ is the probability that category C_i is assigned to a document, without any other information. $P(C_i|t)$ is the probability that category C_i is assigned to a document, given the word t . $P(C_i|\bar{t})$ is the probability that category C_i is assigned to a document, word t doesn't present. $P(t)$ is the probability that word t present in a random document, which $P(\bar{t})$ is the probability that word t doesn't present.

The information gain of term t , $IG(t)$, describes the information we could benefit when classifying documents if we use term t as a feature. And we could find the smaller this score is the less information word t can contribute to the classification algorithm.

We could use information gain as the confusion index to describe the score that word t could confuse the classification algorithm to classify different document classes. For example, if word t 's distributional probability in eight document classes is like the following two scenarios:



For the first scenario, word t 's information gain could be much larger than the second scenario. And we could find the word t , in the first scenario, is already a good feature to identify different classes (we do not need WSD to disambiguate the word senses to train the classifier). However, in the second figure, because the variance between the distributional probability and the average probability is small (with a small information gain), we can hardly classify the documents based on word t (as a feature), and we should use word sense disambiguation to replace word feature with the word sense features. In other words, in the second figure, t is most likely to be shared with different classes, which matches the hypothesis 2 (H2), and should be our focus in this paper.

After we calculate all the possible words' information gain, we could rank them by the score from small to big, and the top n words is the confusion set to be used in the next step.

3.2 Word Sense Disambiguation

From the last step, we get the confusion set of words that need to be disambiguated, and in this step we will find the correct word sense in the context as the new feature to replace the old one. According to previous experience, such as (Miller et al., 1993 & Edmonds et al, 2001), WSD based on human-tagged corpus (supervised learning) achieves the best performance. In this experiment, we will use the SemCor tagged corpus as training data to disambiguate the confusion set of words identified from last section.

The feature types we choose for training follow that of (HT Ng 1996) including: morphological features of word W, part of speech (POS) surrounding W; left and right i-th position POS, unordered set of words frequently co-occur $P(S_i|W_k)$ with W in the same sentence, local collocation, [left offset, right offset] text window, and Syntactic relation, V+N.

Using the above features, we build the context vector for each word and use machine learning to train the model to classify the word senses.

3.3 Clustering word senses

The experimental result of (R Mihalcea 2001 & WB Dolan 1994) show that finding subtle differences in the sense definition could be bad. In the classification task, we won't need to know the difference between senses of bass as a freshwater fish (bass#n#5) and bass as a sea fish (bass#n#6). So the first step is to cluster the similar word senses together.

In this paper we will use the statistical word sense clustering. First, based on the SemCor tagged text, we disambiguate the word (W) (in the confusion set) sense as (S) by using training corpus, which is described in the second step. Based on the distribution of word sense in each document category (C), we could calculate the frequency probability of each word appearing as the specific word sense:

$$P(S_k | W_i, C_j) = \frac{\text{Count}(S_k, W_i, C_j)}{\text{Count}(W_i, C_j)}$$

$\text{Count}(S_k, W_i, C_j)$ is the count of work W_i tagged as the sense S_k in the category C_j . And $\text{Count}(W_i, C_j)$ is the count of work W_i occurs in the category C_j . And if in all the categories:

$$\begin{aligned} \forall \quad & P(S_{k1} | W_i, C_j) > t \\ : \quad & P(S_{k2} | W_i, C_j) > t \end{aligned}$$

where t is the threshold. In this case, we group S_{k1} and S_{k2} together. In other words, category C is not sensitive to the difference between S_{k1} and S_{k2} . In the next step we could make sense S_{k1} and S_{k2} the same feature.

3.4 Document classification

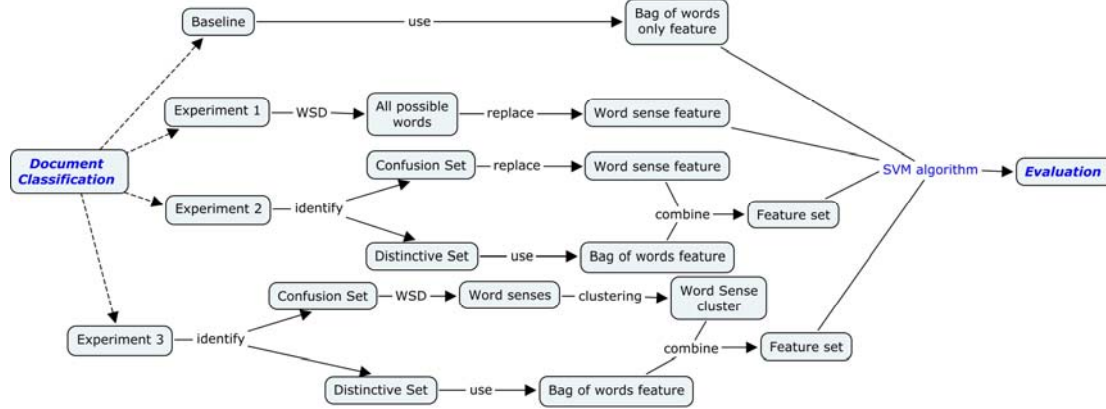
After the three steps above, we add new features derived from word sense clusters. So, the current document vector becomes: W_i ; bag of words (for informative and distinctive word set) + SC_i ; word sense clusters (for confusion set). And we will classify the documents by using the new document vector with SVM (support vector machine) algorithm (Manevitz 2002).

In order to test the hypothesis, we will do several experiments:

- ✓ Baseline: document classification by using only bag of word feature.
- ✓ Experiment 1: Disambiguate all the possible words (without identifying confusion set) and create new word sense features to classify documents
- ✓ Experiment 2: Disambiguate only confusion set of words and create word sense features to classify documents
- ✓ Experiment 3: Disambiguate only confusion set of words, cluster similar word senses together, and use word sense cluster as feature to classify documents

In the Fig2, the particular procedures of methodology are illustrated for the traditional method and new proposal as well. Following the special procedure, man can easily deals with documents in the collection, and each of groups can result in its final evaluation. In the practice, the judgment condition should be followed as below:

- If any of experiment 1, 2 or 3 performance is better than baseline line, hypothesis 1 will be correct.
- If either experiment 2 or 3 is better than experiment 1, hypothesis 2 is correct.
- If experiment 3 is better than experiment 2, hypothesis 3 is correct.



WDS for document classification experiment design

It is very important to consider the influence of several parameters on the final evaluation.

Future work

Instead of focusing on algorithms, future IR and text mining improvements can come from innovative semantic features to represent the documents. In the past decade, researchers have experimented with new features like phrase, name entity, syntagmatic relationships and so on. In this paper, I examine the performance of word sense as a basic semantic unit in the classification task. In fact, we could benefit from word senses much more than to use them as individual units.

In the future, we hope to use them either to group similar word sense features together to create some more powerful feature, or to update the weight of the existing features.

Reference:

- [1] P. Edmonds and S. Cotton. 2001. Senseval-2: Overview. In Proceedings of the Second International Workshop on evaluating Word Sense Disambiguation Systems, Toulouse, France.
- [2] C. Leacock, M. Chodorow, and G. A. Miller. 1998. Using corpus statistics and WordNet relations for sense identification. In Computational Linguistics, volume 24, pages 147–165.
- [3] David Yarowsky, 1995, Unsupervised word sense disambiguation rivaling supervised methods, Proceedings of the 33rd annual meeting on Association for Computational Linguistics, p.189-196
- [4] Yiming Yang, Jan O. Pedersen, A Comparative Study on Feature Selection in Text Categorization, Proceedings of the Fourteenth International Conference on Machine Learning, p.412-420, July 08-12, 1997
- [5] H. T. Ng. 1997. Exemplar-based word sense disambiguation: Some recent improvements. In Proceedings of the Second Conference on Empirical Methods in NLP.
- [6] Xiubo Geng, Tie-Yan Liu, Tao Qin and Hang Li, 2007, Feature Selection for Ranking, SIGIR
- [7] Forman George, 2003, An extensive empirical study of feature selection metrics for text classification, Journal of machine learning research
- [8] David D. Lewis, 1992, An evaluation of phrasal and clustered representations on a text categorization task, Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, p.37-50

- [9] Fabrizio Sebastiani, 2002, Machine learning in automated text categorization, ACM Computing Surveys (CSUR), v.34 n.1, p.1-47
- [10] Viegas, Evelyne, Stephen Beale and Sergei Nirenberg, 1998, The computational lexical semantics of syntagmatic relations, 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics
- [11] Diab, Mona and Philip Resnik, 2002, An Unsupervised Method for Word Sense Tagging using Parallel Corpora, Proceedings of the 40th Annual Meeting on Association for Computational Linguistics
- [12] Rada Mihalcea, 2004, Co-training and Self-training for Word Sense Disambiguation, Proceedings of the Conference on Natural Language Learning
- [13] David Yarowsky, 1993, One sense per collocation. ARPA Workshop
- [14] David Yarowsky, 2000, Hierarchical decision lists for word sense disambiguation, Computers and the Humanities
- [15] William Gale , Kenneth Ward Church and David, 1992, Yarowsky, Estimating upper and lower bounds on the performance of word-sense disambiguation programs, annual meeting on Association for Computational Linguistics
- [16] George A. Miller, 1995, WordNet: a lexical database for English, Communications of the ACM
- [17] William A. Gale, Kenneth W. Church and David Yarowsky, 1992, One sense per discourse, workshop on Speech and Natural Language
- [18] Larry M. Manevitz , Malik Yousef, 2002, One-class SVMs for document classification, The Journal of Machine Learning Research

POSTER SUBMISSION 10:

From Cultural Participation to Information Visualization: A New Framework for Old Knowledge Management Schemas

AUTHOR(S):

Milas, Theodore Patrick

ORGANIZATION(S):

Florida State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The conference theme, “iFutures: Systems, Selves, Society,” calls for research to explore how cultural contexts and information systems affect society. One significant information research area that struggles to address the imperative for cultural sensitivity is knowledge management (KM). Many KM schemas envision meaningful organization of vast knowledge across multiple domains, but ongoing debate about their capacity to do so suggests these attempts fall short. Information research and education turns to other discourse communities, such as epistemology and semiotics, to enrich its approach. To rethink learning for iSchools, this poster presentation proposes innovating KM schemas to reflect archetypes of knowledge systems and information practices manifest in faith-based communities of practice (CoP). Exploring “practice as meaning in particular context,” this qualitative study aimed to balance Etienne Wenger’s duality of reification and participation processes. A content analysis of KM schema qua “hard” knowledge, across religious traditions revealed striking correlations between the conceptual components and spatial representation of knowledge in the mystical symbol systems of two religions commonly considered the most different – Hinduism (polytheism) and Judaism (monotheism). Subsequently, the researcher traveled throughout India, Israel, and Europe to locate communities that believe the schemas of the cakras in tantric Hindu philosophy and the sephirot of Kabbalah (Jewish philosophy) represent ultimate reality. To connect reification of KM with the tacit or “soft” knowledge that informs the everyday lives of religious mystics, the researcher intensively interviewed thirty-six kabbalists and tantrists from a snowball sample acquired via covert

participation in communal worship with the Jewish CoP and by complete observation of the Hindu CoP. Findings suggest ancient KM schemas sustain cultural identity in the research and practices in the Indian and Jewish communities that still use them. In addition to this study's implications for KM, there are also implications for information retrieval: the schemas' central structures are diagrammatically and conceptually identical to Taylor's early model of information needs. Findings suggest that implementation of an integrated system of faceted classification schemas based on the cakras and sephirot should be a new priority in knowledge organization and cultural information systems research and for information visualization applications.

POSTER SUBMISSION 11:

Post-disaster Information Infrastructure: The 1989 Loma Prieta Earthquake

AUTHOR(S):

Finn, Megan

ORGANIZATION(S):

University of California, Berkeley, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Government reports and members of the media blame information systems as the culprit of what has been characterized as faulty responses to disasters such as Virginia Tech and Hurricane Katrina. In the process they propose information and communication technologies (ICTs) as solutions to disaster response. Calls for new ICTs following recent American disasters reveal a poor understanding of the socially situated nature of ICTs. Disaster-related research does not expound a sophisticated understanding of ICTs either. As a consequence, information systems are seen as an unproblematic means of informing the right people with the right information at the right time.

In my dissertation, I will challenge these simplistic notions of information systems as solutions to disaster response by proposing two case studies of "information environments" from the period following the 1989 Loma Prieta earthquake. I will introduce an analytical framework from the social studies of information which situates information systems in the social environment. The theoretical construct, the "information environment," is meant to overcome the limitations of studying ICTs from a purely technical perspective.

The information environment is a synthesis of theoretical work that enables me to look beyond the simple view of ICTs as mere conveyors of information to ask how ICT use is shaped by and shapes the social world. Though the focus of my research is "postdisaster," the information environment construct emphasizes a longer time horizon. For example, one of the information environment approaches poses questions about the values that may be embedded in the design of an ICT. Another approach asks about the conventions of practice around an information infrastructure. Both "values" and "conventions of practice" point to activities which occurred prior to a disaster, calling to mind the question of how far in the past one must look. Additionally, one or both of the "values" or "conventions of practice" concepts might be useful to explain actions following a disaster. The information environment is an intentionally broad (but by no means comprehensive) construct able to include many approaches which socially contextualize ICT use. Part of this project is then determining which aspects of the information environment are most important to understanding ICT use following a disaster. By iteratively going back and forth between the numerous approaches included in the information environment and the case data, I will expand and refine the salient aspects of the information environment construct for the "post-disaster" cases. Thus, this dissertation will not only apply ideas from the social studies of information to expand disaster related research, but also the somewhat novel post-disaster environment will challenge approaches from the social studies of information.

My primary research question is:

(a) How do people use ICTs to make sense of the world in the period following a disaster?

The information environment construct also proposes the questions:

(b) How is ICT use shaped by formal and informal social structures?

My case analysis should also help to answer the question:

(c) How can the information environment be refined and synthesized to best understand ICT use following a disaster?

If I can answer questions (b) and (c), I will gain an understanding of ICTs situated in the local social context. This knowledge will then help to answer my primary question about the cases.

Using the case studies, this dissertation will facilitate an analysis of two propositions:

(i) Understanding the information environment is crucial to discussions about information systems for disaster response.

(ii) Variations in how people use ICTs helps to explain how particular communities can experience the same disaster differently.

In order to address my research questions I propose a theoretically informed, multi-method, pragmatic examination of historical data about the 1989 Loma Prieta earthquake. Using the information environment framework, I will closely examine two cases of community use of an information and communication technology post-disaster. I will look at the following two cases studies: Radio use by the Spanish-speaking community, particularly in Watsonville; and use of the internet, specifically the online community on The Well. Secondary survey data will serve to contextualize my cases within the greater Loma Prieta disaster. By analyzing post-disaster ICT-related practices in two distinct communities, I will show how the information environment framework both elucidates these practices, as well as explains differential post-disaster experiences. The unique combination of historical work on information environments in a post-disaster social context will allow me to make contributions to the social studies of information, and to research on disasters. Finally, my work will provide a case study for doing historical research utilizing new media tools and techniques.

POSTER SUBMISSION 12:

Developing an Organizational Model and Technical Implementation Plan for Wireless Mesh Networks in Sao Tome and Principe, West Africa

AUTHOR(S):

Ritzo, Christopher J

ORGANIZATION(S):

University of Illinois at Urbana-Champaign, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Developing an Organizational Model and Technical Implementation Plan for Wireless Mesh Networks in Sao Tome and Principe, West Africa

Community wireless networking initiatives have been envisioned and implemented with varying success both in the US and abroad. There are a variety of organizational models that municipalities, community organizations, individuals, businesses, corporations and grassroots movements are using to build, plan and implement wireless community networks. In developing countries the organizational, economic and technical challenges are even greater. Terrain, power requirements, and other technical challenges are secondary to the organizational challenges, community impact and social implications that such a technology infusion can bring. These needs present a great opportunity for Universities to complete research, help communities bridge the digital divide, and become part of the information society.

Prairienet, a Community Informatics focused organization within the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign, has been partnering on action research projects with organizations and people in communities such as East St. Louis, IL and Paseo Boricua, Chicago for a many years. This has provided graduate students in Library and Information Science with opportunities to learn valuable technical and organizational skills, while assisting community groups achieving their community goals.

Prairienet is now beginning to extend this model internationally through partnerships begun through former GSLIS graduate, Jorge Coelho of Sao Tome and Principe. Initial discussions on community goals with Sao Tomean community members and groups over the past two years have developed into several projects carried out by both GSLIS graduate students and those from the Architecture and Urban Planning schools. These early projects have benefited both Sao Tomean communities and the students engaged with the community members completing them.

An early expressed goal of Jorge and other community members has been the desire to link locations with Internet access such as schools and libraries with community centers in nearby plantation towns in the mountains. Initial surveying of two possible sites and partial testing of open source mesh network hardware and software took place as a small part of Prairienet's summer 2007 trip. In response to the lessons learned from this testing, GSLIS student Chris Ritzo has begun work investigating the different organizational models being used in community wireless network initiatives, continued testing open-source mesh networking technologies in use or in development in the U.S. or internationally and comparing those models and technologies to organizational and personal contacts of the GSLIS/Prairienet Community Informatics initiative in Sao Tome, West Africa. This research will result in a paper detailing possible organizational and technical models for deployment of wireless mesh networks in Sao Tome and Principe. This might then be used by Sao Tomean community members and organizations in partnership with the University of Illinois/GSLIS groups to make that goal a reality, or as a starting point for further discussion and modification.

This poster session will detail the work completed thus far, discussing possible organizational and technical models that could be used to implement wireless networking initiatives in Sao Tome and Principe.

POSTER SUBMISSION 13:

Deriving Ontology from Folksonomy and Controlled Vocabulary

AUTHOR(S):

Chen, Miao

Qin, Jian

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Deriving Ontology from Folksonomy and Controlled Vocabulary

Introduction

Popular online tagging websites, such as Flickr, Technorati, and Del.icio.us, allow users to tag objects freely without constraints of any controlled vocabulary. The word “folksonomy” has been used to describe this type of grass-root taxonomies, which provides a rich source for building ontologies.

Research has experimented with building ontology purely from user-generated tags, and the approaches have been focused primarily on statistical methods. Schmitz (2006) conducted a study of inducing an ontology from Flickr tags, which used statistical methods to detect subsumption pairs based on co-occurrences of tags. Heymann Carcia-Molina (2006) established a hierarchical taxonomy based on tags from Delicious and CiteULike respectively by using cosine similarity of tag vectors.

Although a good source for building ontology, folksonomy has its disadvantage in representing object content. For example, the tags are subjective and their hyponyms are usually not indicated (Mathes, 2004). Relations between tags are unknown and useful information is missed in tags, therefore the unorganized status of tags affects retrieval of objects negatively. Folksonomies alone are not sufficient for building a comprehensive and high-quality ontology. Researchers have suggested using other sources such as WordNet and Wikipedia to assist ontology building from folksonomies (Damme et al., 2007).

Contrary to folksonomies, controlled vocabulary is characterized by rigid structures and slow responsiveness to new terminology. But its systematic organization and careful formulation of terms and relationships would be complementary to the disadvantages of folksonomies. Online lexical resources and gazetteers as instances of controlled vocabulary appear to be promising approaches in using folksonomies for generating ontologies (Schmitz, 2006; Damme et al., 2007). While researchers are speculating this method, little research has been done to actually implement it to build ontology. In addition, building ontologies from a combination of both folksonomies and controlled vocabulary is rarely mentioned in previous works. This study is intended to fill the gaps between methodologies in using folksonomies to produce ontologies.

Assumption

Tags are keywords used to index objects and one object (such as a photo or a webpage) may be tagged by one or more tags. Each tag may be associated with other tags through co-occurrences. We assume that frequent co-occurrences do not happen by accident in statistical terms. If one tag frequently co-occurs with another tag, then there should be some relationship between them. The co-occurring tags of one tag are called “related tags”. We emphasize related tags in this study because related tags can offer useful semantic information about tag relationships.

Methodology

This study uses an approach of combining user-generated tags and controlled vocabulary to develop an ontology on landscape. The controlled vocabulary is used as the backbone (Mani et al., 2004) for the ontology. We then expand the knowledge structure by adding more entries from user-generated tags. The existing terms in the controlled vocabulary serve as classes in the ontology, with clear hierarchy of classes and subclasses (Qin & Paling, 2001). Based on the tags, we build a micro-hierarchical system for each class. The micro-hierarchies may be deployed in two ways: one is used as updates for the controlled vocabulary and the other is used as assemblages to form a portable ontology. This is the overview of the method, and the method of building hierarchy for subclasses is presented below.

Tags can be represented by their related tags. For a tag T , its related tags are rt_1, rt_2, \dots , and rt_n , which forms a vector for the tag $T(rt_1, rt_2, \dots, rt_n)$. All the tags in the domain (the subclass domain) can be converted to the vector format. Then we further build a hierarchy for all the tags based on the vector similarity scores. Clustering techniques is applied in extracting hierarchy from the tags. We compute the vector similarity between tags and cluster the tags to form a hierarchy. Relations will be added to describe the relations between classes of the hierarchy manually.

Data and Implementation

For the pilot study, we chose the field of landscape images as the ontology domain, and used Flickr tags and Alexander Digital Library (ADL) geographic feature type thesaurus as controlled vocabulary. Flickr users tag photos with keywords and the system has cumulated more than a million photos tagged by "landscape." The ADL thesaurus was extracted into a database and the level of each term in the hierarchy is indicated. For example, "gulfs" is labeled level 2, with its upper category "hydrographic features" as level 1.

The tag data were collected from the related tags from the Flickr API `flickr.tags.getRelated`. For example, "sky" is one of the related tags of "landscape," and "landscape" can be represented by vector (nature, sky, clouds, trees, ..., evening). Figure 1 demonstrates how the "lakes" subclass of ADL thesaurus is expanded into a hierarchy.

The data set of tags is built upon tag "lakes" by collecting the related tags, related tags of the related tags, and so on, until the number of tags reached the threshold 100. The data set can be viewed as tags in the small domain of lake. With the related tags provided by Flickr API, we can obtain vector representation of each tag in our database. Clustering is applied on the vectors to explore hierarchy among the tags. With the updated hierarchy, relations are added between classes. The workflow and mock-up result are shown in Figure 1.

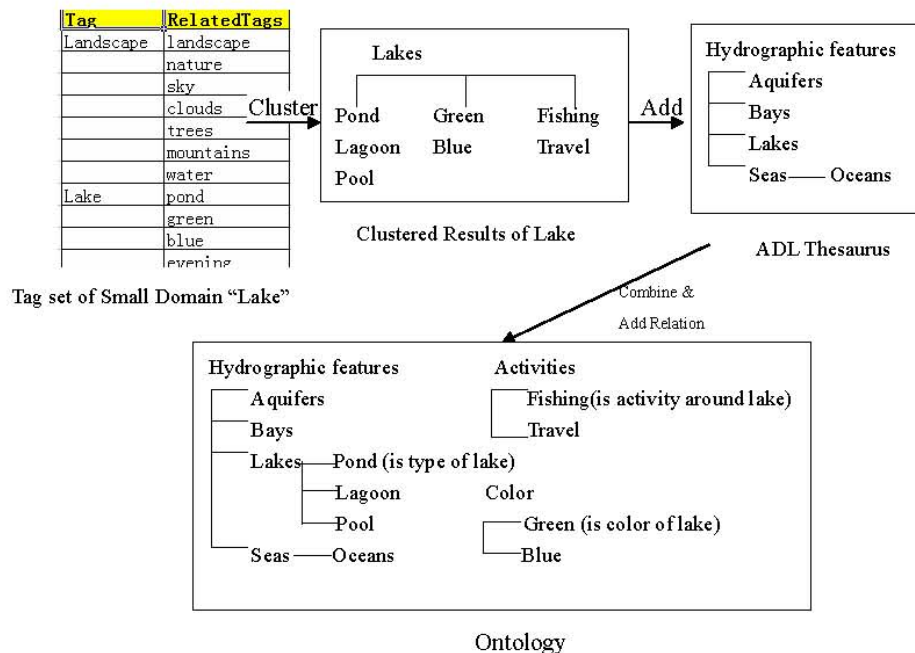


Figure 1. A mock-up result of ontology generation from tags

Future Plan

Knowledge capture is the bottleneck problem in intelligent information systems. Developing ontologies from user-generated tags in combination of controlled vocabulary is our contribution to solving this bottleneck problem.

In this poster proposal, we proposed a method of building ontology from folksonomies and controlled vocabulary. A pilot study was conducted in the field of landscape with partial completion. The next phase of research is to create a training data set by building knowledge base that contains tags, controlled vocabulary, and reasoning rules for automatic ontology generation and relation assignment. While the training data set involves manual work due to its complexity and lack of prior data in this regard (Qin & Paling, 2001; Ding & Foo, 2002), we expect to automate this process in future research.

References

- Alexandria Digital Library Feature Type Thesaurus. Retrieved October 29, 2007 from <http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/index.htm>
- Chung, C.Y., Lieu, R., Liu, J., Luk, A., Mao, J., & Raghavan, P. (2002). Thematic mapping-From unstructured documents to taxonomies. CIKM '02, Virginia, USA.
- Damme, V.C., Hepp, M., & Siropaces, K. (2007). FolksOntology: An integrated approach for turning folksonomies into ontologies. ESWC 2007 "Bridging the Gap between Semantic Web and Web 2.0" workshop.
- Ding, Y., & Foo, S. (2002). Ontology research and development. Part1-A review of ontology generation. *Journal of Information Science*, 28(2), 123-136.
- Heymann, P., & Carcia-Molina, H. (2006). Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical Report 2006-10, Department of Computer Science, Stanford University.
- Mani, I., Samuel, K., Concepcion, K., & Vogel, D. (2004). Automatically inducing ontology from corpora. 3rd International Workshop on Computational Terminology, COLING'2004, Geneva.
- Mathes, A. (2004). Folksonomies-Cooperative classification and communication through shared metadata. Retrieved October 29, 2007, from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Qin, J., & Paling, S. (2001). Converting a controlled vocabulary into an ontology: The case of GEM. *Information Research*, 6(2). Retrieved October 29, 2007, from <http://informationr.net/ir/6-2/paper94.html>
- Schmitz, P. (2006). Inducing Ontology from Flickr Tags. WWW 2006, Edinburgh, UK.

POSTER SUBMISSION 14:

A visual map of public mementos and conjectures

AUTHOR(S):

Pepe, Alberto

ORGANIZATION(S):

University of California, Los Angeles, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

How do we collectively feel about our future? Do we look forward to it with anxiety or vigor? Are we apprehensive or optimistic of what the future will bring? Since mood affects performance and well-being, the answers to these questions matter greatly to anyone concerned with public policy. The web is awash with material indicative of public mood, collective forecasting and personal relics. Several efforts have been undertaken to assess emotional status from online sources such as blogs, emails, web sites (Balog & De Rijke 2006) and search engine queries (see, for example, Google Trends). However, these efforts are limited, by the nature of their source material, to hindsight and near-present observations.

The work presented here is concerned with collective speculations about the future. We present a visual analysis of publicly available textual content from futureme.org, a popular web service that allows its users to send themselves emails to be delivered at a later date, up to 30 years in the future. Many of these emails resemble "confessional" time capsules: their content is intended to project the user's present emotional state at the origination date towards the indicated delivery date. These emails fall into two broad categories of content: a) conjectures about the future and b) mementos regarding the present or the past. By aggregating mood indicators extracted from messages directed to future dates, we can thus assess short and long term shifts in the collective emotional perception toward a particular point in the future. This principle is related to "wisdom of crowd" phenomena as observed in finance and prediction markets (Surowiecki 2004).

Numerous psychometric instruments to assess individual mood states and monitor their fluctuations over time exist, the most prominent of which is the 65 item Profile of Mood States (POMS) questionnaire (McNair, Lorr, & Droppleman 1971). The 6 dimensional factor analytical structure of the POMS (tension, depression, anger, vigor, fatigue and confusion) has been validated repeatedly (Norcross, Guadagnoli, & Prochaska 2006)

and applied in hundreds of studies since its inception (McNair, Heuchert, & Shilony 2003). To make the POMS questionnaire applicable to the open-ended nature of email content, we extended the POMS set of 65 adjectives by nearly 793 synonyms using WordNet and Roget's Thesaurus.

We calculated the occurrences of extended POMS terms in the content of 30,000 publicly available "future" emails and mapped them to a normalized six-dimensional mood vector representing levels of tension, depression, anger, vigor, fatigue and confusion. These mood vectors were grouped according to the delivery date of the original email, resulting in a set of mood state vectors. Statistically significant mood changes were detected especially for depression and vigor indicators. The computation of mood levels was then implemented with a more specific textual analysis of the entire email corpus, aimed at identifying manifestations of conjectures and mementos.

The results, presented in this poster, blend two different visual representation of the content analyzed: an "emotional timeline" - a cumulative depiction of mood levels between 2007 and 2036 - and a superimposed topic map of mementos and conjectures - an ontological model of commonly used terms and adjectives illustrating the chains of word association.

References

Balog, K., and De Rijke, M. 2006. Decomposing bloggers' moods. In World Wide Web Conference.

McNair, D.; Heuchert, J. P.; and Shilony, E. 2003. Profile of mood states. Bibliography 1964–2002. Multi-Health Systems.

McNair, D.; Lorr, M.; and Droppleman, L. 1971. Profile of mood states.

Norcross, J. C.; Guadagnoli, E.; and Prochaska, J. O. 2006. Factor structure of the profile of mood states (POMS): Two partial replications. *Journal of Clinical Psychology* 40(5):1270 – 1277.

Surowiecki, J. 2004. *The Wisdom of Crowds*. Doubleday.

POSTER SUBMISSION 15:

Group Maintenance Behaviors in Self-Organizing Distributed Teams

AUTHOR(S):

Li, Na (Lina)

Scialdone, Michael John

Heckman, Robert

Crowston, Kevin

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Group Maintenance Behaviors in Self-Organizing Distributed Teams

(Abstract of a Poster submission)

Introduction

Scientific, industrial, and other corporate work is increasingly being conducted in self-organizing distributed teams enabled by information and communication technologies (ICTs). However, the limitations of ICTs and the geographical, organizational and social distance between distributed group members challenge them to maintain social relationships necessary for group effectiveness. Understanding group maintenance may provide insight into the success or failure of such teams. Group maintenance is defined as discretionary, relation-building behavior that enables group members to trust and cooperate with one another more easily [1]. This behavior has been found to affect group outputs such as team effectiveness and member satisfaction [2].

This study examines group maintenance behavior in Free/Libre Open Source Software (FLOSS) development teams as examples of distributed teams. Most FLOSS software is developed by dynamic, self-organizing distributed teams of professionals, users and other volunteers working in a loosely coupled manner [3-5]. These teams are close to pure virtual teams in that developers contribute from around the world, meet face-to-face infrequently if at all, and coordinate their activity primarily by means of computer-mediated communications (CMC) [6, 7].

Literature review

Group maintenance behaviors are those that contribute to the creation of an environment that supports a work group's task-related activities. Based on literatures in social presence [8, 9], social-emotional behavior [10], politeness theory [11-13], and organizational citizenship behavior (OCB) [14, 15], we developed a coding scheme to identify group maintenance behaviors in virtual teams. The scheme, shown in Appendix Table 1, has four top-level categories: emotional expression, positive politeness, negative politeness, and OCB. Emotional expression refers to conventional and unconventional expressions of emotion. Positive politeness tactics are those that help group members locate common ground and bond as a group, whereas negative politeness tactics show respect for the autonomy of others and maintain appropriate distance [16]. Examples of positive politeness could be use of inclusive pronouns, expression of agreement and complementation. Use of self-depreciation, verbal hedges, and formal verbiages are good examples of negative politeness. Organizational citizenship behavior refers to "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and in the aggregate promotes the efficient and effective functioning of the organization" [14]. Helping and courtesy behaviors are OCB examples.

Method

This study employs a multiple case study method. We chose two FLOSS projects that developed Instant Messaging (IM) clients: Gaim and Fire. The two projects were similar in terms of their project goals, nature of tasks, and potential users. However Gaim has been more effective as a project, based on Crowston et al's multivariate measure of effectiveness in FLOSS contexts [17, 18].

The nature of FLOSS teams provides unique opportunities to observe group maintenance behavior since most FLOSS activities (if not all) are archived. We collected messages sent to the developers' email lists or forums by all contributors. We differentiated three stages in both projects. At each stage, twenty episodes were identified based on topics in discussion and attention from members. We took the first 20 episodes available as the beginning stage and last 20 as the ending stage. The middle stage is located around a major software release approximately halfway between the beginning and ending stages. Each episode consisted of multiple email messages focusing on a particular issue. A total of 60 episodes are under investigation for each project.

We conducted content analysis to examine group maintenance behavior in the two projects. We adopted a thematic measure as our unit of analysis: "a single thought unit or idea unit that conveys a single item of information extracted from a segment of content" or the "unit of meaning" [19]. Such units vary in size from an emoticon or punctuation to a word, a phrase, a part of a sentence, a sentence, or even a few sentences when appropriate. The coding scheme was initially created deductively from the literature reviewed above. We then revised it according to a small number of episodes in both projects. The revised scheme was then used to code more episodes and revised again. The iteration process repeated until a relatively solid coding scheme was achieved (see the appendix). Two coders coded the episodes independently with an inter-rater reliability of 0.79. The two coders then discussed disagreements to reach consensus.

Result & Discussion

In this abstract, we report on 48 of the 120 episodes that have been fully resolved by the coders (24 for each project). The 24 Gaim episodes were consisted of 214 messages, receiving a total of 2230 group maintenance codes; while the 24 Fire episodes included 164 messages, to which 1756 codes were assigned. Usually, episodes containing more messages tend to reveal more group maintenance behaviors, so comparison between the groups on frequencies may be misleading. Therefore we calculated the densities of group maintenance behaviors. We define density as the number of codes in a unit/number of messages in a unit. Density can be calculated for an episode, a stage, or the whole project. Table 1 in the appendix shows the densities of each code and each category in both groups. The average densities across 4 categories of group maintenance behaviors are 10.42 and 10.71 for Gaim and Fire, respectively.

Positive politeness behaviors have been widely observed in the two projects. The densities are 7.14 and 7.63 for Gaim and Fire respectively. Specifically, group-specific jargon/metaphor, colloquialism/slang, inclusive pronoun, and vocative appeared most frequently. This suggests that members view their team as a cohesive group that has its own culture that bonds its members together.

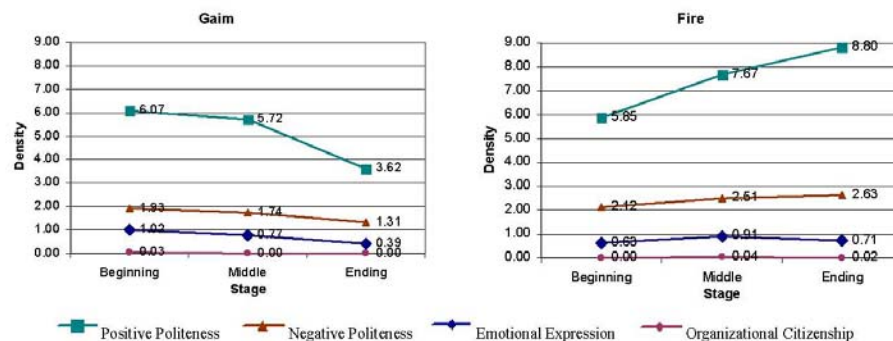
Negative politeness tactics densities for Gaim and Fire are 2.23 and 2.38 respectively. The most dense code found was hedges/hesitation/subjunctives (e.g., "I'm not sure...", "it would be nice if ..."), suggesting that team members are somewhat careful to respect the autonomy of others despite the casual atmosphere.

Emotional expressions have been employed to a moderate extent with densities for Gaim and Fire at 0.32 and 0.37 respectively. Both groups used punctuation most often to express emotions or emphasis. However, Gaim employed the other 4 types of emotional expressions (emoticon, capitalization, explicit emotion, and humor) more than Fire did.

On the other hand, organizational citizenship behaviors (OCB) were very rarely seen, with an average density of only 0.02 for both projects. The low density of these behaviors may reflect the fact that OCB theory was developed in traditional organizations, where member responsibilities are clearly defined. Under this condition, it is easy to identify behaviors such as helping or courtesy that go beyond what is required. In the FLOSS setting, however, most members contribute to the projects voluntarily and it is hard to distinguish the line between fulfilling one's own responsibility and helping others.

Figure 1 shows the change of emotional expression, positive politeness, negative politeness, and OCB over time by stage in both projects (the lines serve only to visually link related points). The figure demonstrates that, as the projects proceeded, all 4 categories of behaviors were decreasing in Gaim; while positive and negative politeness behaviors went up in Fire. This finding is opposite to our expectation as Gaim was the more successful project as noted above. One possibility is that the more successful team developed a friendlier atmosphere and higher level of participation earlier, so that less relationship maintenance behaviors were needed later; while the less successful team was still struggling with member relationship and recruitment until the end of the project. Another possibility could be increased usage of instant messaging tools over time in Gaim with group maintenance discussions migrating away from email.

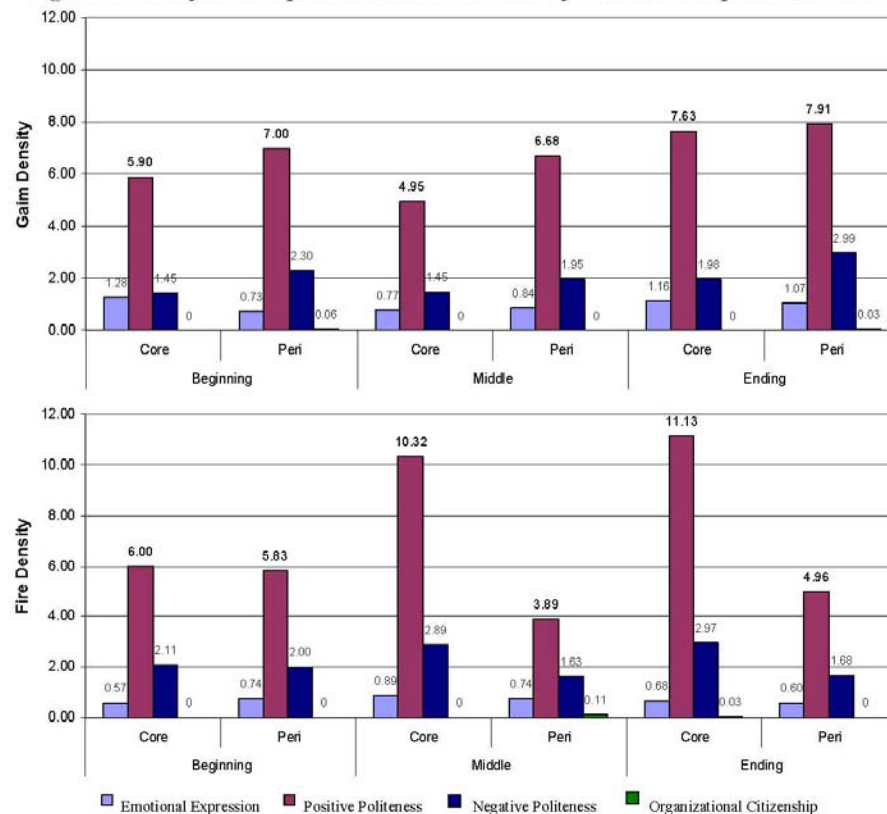
Figure 1. Density of Group Maintenance Behaviors by Stage in Gaim and Fire



We are interested in whether there was a difference in the use of group maintenance behaviors between core members (defined as project administrators and developers as indicated by status on the SourceForge system) and peripheral members (normal users). Figure 2 shows the densities of 4 types of behaviors employed by core and peripheral members at various stages in both projects. Fire started with both types of members performing group maintenance behaviors with similar densities. In the middle and ending stages, however, core members employed positive and negative politeness tactics in a much denser manner than peripheral members. In Gaim,

interestingly, we observed not much difference between core and peripheral members. Indeed, at all stages, peripheral members showed slightly denser behaviors than core members on OCB, positive and negative politeness. Meanwhile, core members expressed emotions a little more than peripheral members at the beginning and ending stages.

Figure 2. Density of Group Maintenance Behaviors by Core and Peripheral Members



Future Research

Our plans for completing this work include finalizing the coding for the remaining episodes, as well as coding additional projects. We will conduct t-test and ANOVA to examine the difference between groups and within group (between core and peripheral members, among different stages). We also plan to investigate the relationship between group maintenance behaviors and the maintenance of membership and team effectiveness. Another step is to investigate how individual core and peripheral members change their way of conducting group maintenance behaviors over time.

Appendix.

Table 1. Coding Scheme of Group Maintenance Behavior

Category	Indicators	Definition	Example	Gain Density	Fire Density
Emotional Expression	Expressions of emotion using emoticons.	Expressions of emotion or emphasis using emoticons.	:)	0.16	0.08
	Expressions of emotion or emphasis using conspicuous capitalization.	Expressions of emotion or emphasis using conspicuous capitalization.	"EVERYONE ON THE LIST" "AND" "THINK"	0.16	0.10
	Expressions of emotion or emphasis using (repetitious) punctuation	Expressions of emotion or emphasis using (repetitious) punctuation, exclamation point, underlining, italic fonts, or any other	"!!!" Underline "!"	0.32	0.37
	Explicit expression of emotion	Direct or explicit expression of emotion using emotional words. (we need to refresh ourselves to the hundred words at this moment)	"How exciting! The first post!"	0.17	0.11
	Use of humor	Teasing, cajoling, irony, understatements, sarcasm. So we will code every repeated signature, but not coded text.	"it also tells you how many twinkies it takes to run this mailing list. It's a lot of twinkies." "The only way to keep your health is to eat what you don't want, drink what you don't like, and do what you'd rather not. -- Mark Twain"	0.21	0.08
Positive Politeness	Colloquialisms or slang	Spelling out phonological slurring, using colloquialisms or slang; beyond group specific; used to show familiarity.	"Saturdayish" "yep" "BTW"	0.56	0.36
	Group-specific jargon or metaphors	Use of group-specific jargon, language, or metaphors.	"Why is this a .mm file? what is .mm again? I know .m is ObjC"	4.65	5.40
	Vocatives	Referring to participants by name, or specifically addressing part of a message to an individual. Name used as the second or third or even first person. If there is a "you" or "your" specifically referring to a particular single person, we'll code it.	"As sean said" "Martin,"	0.55	0.56
	Inclusive pronouns	Incorporating writer and recipient(s)	"we"	0.69	0.78

Category	Indicators	Definition	Example	Gaim Density	Fire Density
			"us" "let's" "our"		
	Phatics	Personal greetings and closures, including communication for purely social reasons	"Hi" "regards" "Thanks for the help." (at the end of a message)	0.22	0.12
	Raising/presupposing commonalities	Assuming a stance that is in agreement with other group members – attributes of things that we share.	"If you have two patches that modify the same file, it is best to separate them. I know this is a pain, but I go through this every time I submit something to the gaim guys since we use their library."	0.01	0.01
	Expressing empathy/sympathy	Expressing empathy and/or understanding	"I know this is a pain"	0.02	0.01
	Self-disclosure	Providing information specifically on self, opening up to the group (shouldn't be coded as explanation"; if needed, we can assign it multiple codes)	"I am not working actively on Fire due to personal and professional issues right now."	0.05	0.02
	Complimenting	Complimenting others or message content. It can go beyond agreement.	"The temporary message is a good idea"	0.06	0.04
	Expressing agreement	Expressing agreement with others previous statement	"Agreed" "I suppose." "Correct."	0.14	0.11
	Admitting Mistake/Fault	Apologizing for one's own personal mistakes	"Sorry again if I stepped on any toes"	0.01	0.01
	Encouraging participation	Encouraging all the members of the group to participate	"Any comments welcome.";	0.10	0.12
	Expressing appreciation		"Thanks for the help."	0.07	0.08
Negative Politeness	Disclaimers, Self-depreciation	Use of disclaimers prior to an FTA; self-depreciation as a distancing tool; may include apologies as explanations	"dumb fire question#1: which MSNService nib "file" is the real one?" "Sorry if I'm terribly ignorant somehow... I'm just getting into this stuff."	0.24	0.22
	Stating rational for	Stating an FTA as a general rule to minimize impact or	"In general we want to avoid forking the	0.18	0.14

Category	Indicators	Definition	Example	Gain Density	Fire Density
	FTA (State FTA as general rule + explanation)	as to not single out an individual; Explaining the reasons behind an action that might threaten someone's face.	MSN library with our own changes so any changes there need to be sent on to Meredydd. "Since this history thing requires a bunch of functions I created history.c."		
	Indirect inquiry	Using a phrase to diminish the force of inquiry. Inquiring into hearer's ability/willingness to comply through indirect speech (ie. Dropping hints)	"Is there nay chance you could apply my patch to the current version though?" "If you (see if this "you" is a specific person) want to take a look that would be great."	0.04	0.13
	Hedges; Hesitation; subjunctives	Use of words/phrases/subjunctives to diminish force of act; Use of hesitation in disagreement (ie. "well...")	"um..." "I'm not sure what the problem is..." "it would be nice to at least..."	1.63	1.77
	Honorifics	Use of honorifics	Mr., Miss., Dr., Prof., etc.	0.00	0.00
	Formal verbiage	Using formal wording choices	"please send the file to..."	0.03	0.04
	Impersonalization	Avoid use of I or you to diminish the force of inquiry or action in cases where one would normally have to use you or I	"Can anyone point to any available docs. & how to start on it?"	0.05	0.02
	Passive voice used to create distance	Use of passive voice to create distance between the object from the performer	"What is it that isn't handled properly?" "Is it being worked on?"	0.07	0.05
Organizational Citizenship Behaviors	Helping	Behavior involving voluntarily helping others with a work problem. The immediate beneficiary is a specific individual person.	"I'll look into that."	0.00	0.01
	Courtesy	Subsumes all of those foresightful gestures that help someone else prevent a problem; avoiding practices that make other people's work harder.	"Note that, after applying this patch the SILC Toolkit 1.0.1 must be installed in the system in order to be able to compile."	0.00	0.00
	Peacemaking	Actions that help to prevent, resolve, or mitigate unconstructive interpersonal conflict.	"A compromise could be to..."	0.00	0.01
	Cheerleading	The words and gestures of encouragement and reinforcement of coworkers' accomplishments and professional development.	"You will be honored for contributing your time and skill to a worthy cause."	0.01	0.01
	Sportsmanship	A willingness to tolerate the inevitable inconvenience and impositions of work without complaining.		0.00	0.00

References

- [1] M. Ridley, *The Origins of Virtue: Human Instincts and the Evolution of Cooperation*. New York: Viking, 1996.
- [2] J. R. Hackman, "The design of work teams," in *The Handbook of Organizational Behavior*, J. W. Lorsch, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1987, pp. 315–342.
- [3] E. von Hippel, "Innovation by user communities: Learning from open-source software," *Sloan Management Review*, pp. 82–86, 2001.
- [4] E. von Hippel and G. von Krogh, *Exploring the Open Source Software Phenomenon: Issues for Organization Science* Sloan School of Management, MIT, 2002.
- [5] E. von Hippel and G. von Krogh, "Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science," *Organization Science*, vol. 14, pp. 209–213, 2003.
- [6] P. Wayner, *Free For All*. New York: HarperCollins, 2000.
- [7] E. S. Raymond, "The cathedral and the bazaar," *First Monday*, vol. 3, 1998.
- [8] R. Garrison, T. Anderson, and W. Archer, "Critical thinking in a text-based environment: Computer conferencing in higher education," *The Internet and Higher Education*, vol. 2, pp. 87–105, 2000.
- [9] L. Rourke, T. Anderson, D. R. Garrison, and W. Archer, "Assessing social presence in asynchronous, text-based computer conferencing," *Journal of Distance Education*, vol. 14, pp. 51–70, 1999.
- [10] D. S. Gouran, C. Brown, and D. R. Henry, "Behavioral correlates of perceptions of quality in decision-making discussions," in *Communication Monographs*, 1978, pp. 51–63.
- [11] D. S. Gouran, C. Brown, and D. R. Henry, "Behavioral correlates of perceptions of quality in decision-making discussions," *Communication Monographs*, pp. 51–63, 1978.
- [12] R. J. Ocker, "The Relationship between Interaction, Group Development, and Outcome: A Study of Virtual Communication," in *Proceedings of the Thirty-fourth Annual Hawaii International Conference on System Science*, 2001.
- [13] D. A. Morand and R. J. Ocker, "Politeness theory and Computer-Mediated communication: A Sociolinguistic Approach to Analyzing Relational Messages," in *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*, 2003.
- [14] D. W. Organ, P. M. Podsakoff, and S. B. MacKenzie, *Organizational citizenship behavior: Its nature, antecedents, and consequences*. Thousand Oaks, CA: SAGE Publications, 2006.
- [15] P. M. Podsakoff, S. B. MacKenzie, R. H. Mooreman, and R. Fetter, "Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors," *Leadership Quarterly*, vol. 1, pp. 107–142, 1990.
- [16] D. A. Morand and R. J. Ocker, "Politeness theory and Computer-Mediated communication: A Sociolinguistic Approach to Analyzing Relational Messages,"

- in *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS)*, 2003.
- [17] R. Heckman, K. Crowston, Q. Li, E. Allen, U. Y. Eseryel, J. Howison, and K. Wei, "Emergent decision-making practices in technology-supported self-organizing distributed teams," in *Proceedings of International Conference on Information Systems*, 2006.
 - [18] K. Crowston, J. Howison, and H. Annabi, "Information systems success in Free and Open Source Software development: Theory and measures," *Software Process - Improvement and Practice*, vol. 11, pp. 123-148, 2006.
 - [19] R. Budd, R. K. Thorpe, and L. Donohue, *Content analysis of communication*. New York: Macmillan, 1967.

POSTER SUBMISSION 16:

To Catch a Predator? The MySpace Moral Panic

AUTHOR(S):
Marwick, Alice E.

ORGANIZATION(S):

New York University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

This paper is about moral panics over contemporary technology, which I call “technopanics.” I use two examples, the cyberporn panic of 1996 and the contemporary panic over online predators and MySpace, to demonstrate the links between media coverage and content legislation. In both cases, Internet content legislation is directly linked to media-fueled moral panics that concern uses of technology deemed harmful to children. This is of particular interest right now as an internet content bill known as the Deleting Online Predators Act, or DOPA, was debated last year in Congress, and a variant of this bill may pass in the future. DOPA would require filtering software on internet-connected computers in libraries and schools to block access to all “social software” sites, which would broadly include any website with communication features, including social networking sites, wikis, community forums, and online communities. The technopanic over “online predators” is remarkably similar to the cyberporn panic; both are fueled by media coverage, both rely on the idea of harm to children as the justification for internet content restriction, and both have resulted in carefully crafted legislation to circumvent First Amendment concerns. While both panics have their roots in legitimate concerns, my perspective is not that of a realist – I am not primarily concerned with the extent of the purported harms. However, my research demonstrates that the legislation proposed (or passed) to curb these problems is an extraordinary response; it is misguided and in many cases masks the underlying problem.

The first part of this paper discusses the cyberporn panic of the mid-1990s and its links to internet content legislation. Second, I outline a theory of technopanics, using the moral panics literature to examine fears of technology and modernity. The third section examines the current moral panic over social networking sites and MySpace.com in particular, including the DOPA legislation proposed to solve the problem. The fourth part of this paper takes a critical look at the harms of MySpace and online predators in general. I argue that in order to justify the passage of sweeping laws restricting all minors’ access to MySpace, the problem of online predators using MySpace to elicit underage individuals for sexual activity has been portrayed as socially significant, which I do not believe to be the case. Rather, I believe that it is indicative of a small, visible number of sensationalized cases that have been used by the media to fuel a moral panic around the site. The idea of the “online predator” is further linked to real harms, such as child abuse and child pornography, to create rhetorical significance for legislative-based action. I conclude with some recommendations for legislators, parents, and journalists.

Since moral panics are, by definition, public, any study of a moral panic must attempt to access popular discourse. This paper primarily uses Congressional testimony and popular press articles, both national and local, as evidence. Congressional testimony was found by conducting full-text searches of the Congressional Record, a publication which includes every word spoken on the floor of both the House and the Senate. Popular press articles were found through Lexis-Nexis, the websites of major and local newspapers, and television network and cable news programs.

POSTER SUBMISSION 17:

Longitudinal Analysis of Freshman Adoption of Facebook.com

AUTHOR(S):

Stutzman, Frederic

ORGANIZATION(S):

University of North Carolina-Chapel Hill, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

**iConference 2008
Poster Submission**

Poster Title: Longitudinal Analysis of Freshman Adoption of Facebook.com

Abstract: Originally designed as a social network for college students, Facebook¹ is a leading social network site at American colleges and universities, with high degrees of adoption and market penetration (Gross and Acquisti, 2006). In the following poster, results of a 16-week longitudinal study of freshman adoption of Facebook.com at [X University] will be presented.

Description: This study examines adoption behavior and use of Facebook at [X University] over the length of Fall semester, 2005 (August-December, 2005). Using custom-designed crawler software, data on the freshman cohort was collected weekly from Facebook, capturing the growth of the social network over the freshman semester.

College students rely on social networks for various reasons. Primarily, students create digital representations of identity through the enumeration of a social network profile. Data in the profile may include elements like interests and hobbies, jobs or relationship status. This data serves a variety of purposes; it establishes the student as individual in the network, it opens the student to new interactions, and it provides an anchor for contact and discovery over time. Described by boyd, the creation of the profile writes the individual into being in the eyes of the community (boyd, 2006).

The profile also serves an important consumptive purpose on the college campus. Residential students often leave social networks behind as they immerse themselves in university; this period of intense transition creates unique social needs (Hays and Oxley, 1986). For first-term freshmen, the negotiation of new social networks is a primary need; students turn to social network sites to help them in this negotiation. Research by Lampe, Ellison and Steinfeld (2006) found that college relationships in Facebook move offline-to-online, signifying the value of the social network profile as an information point to be explored and consumed by potential friends.

Beyond construction and consumption of identity information, students also enumerate social network connections through “friending” – the process of explicitly stating connection in a social network site. Of course, a friendship in a social network site is not equivalent to a “real world” friendship, but rather a “weak tie” representative of connection and relation. Acknowledging this reality, the structure and evolution of a college social network as realized in this dataset remains remarkable.

¹ <http://facebook.com>

Analysis of this dataset (n_1 3196, n_{16} 3524) will focus primarily on profile construction and friending behavior over the sixteen weeks of data collection. Preliminary analysis of the data reveals interesting information about date of profile creation (Figure 1), Facebook participation (Figure 2) and social network growth (Figure 3).

In this preliminary analysis, I asked three simple questions of the dataset. The first explored when 2005 freshman created the Facebook profiles.

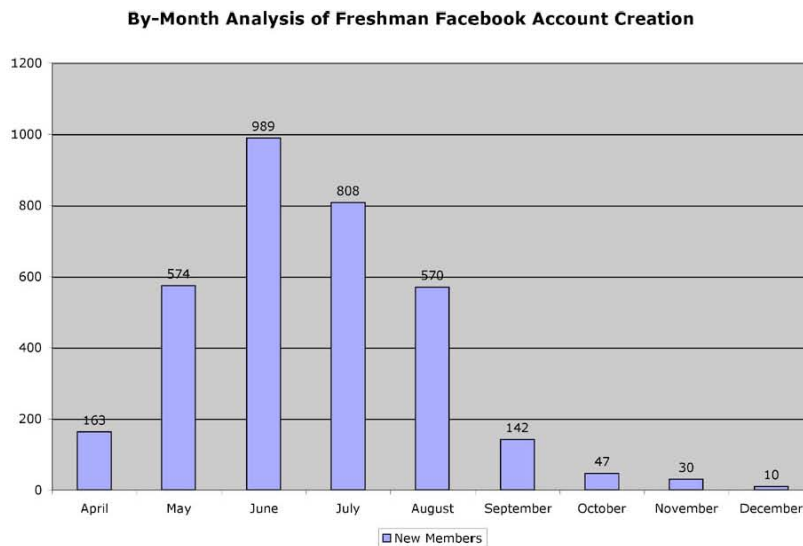


Figure 1. By-Month Analysis of freshman Facebook Account Creation

While school began in August, one can clearly see the majority of Facebook profiles were created over the summer. Exploring this data, I am able to see that profile creation correlates to orientation weekends, indicating the freshmen were likely learning about Facebook via word-of-mouth at orientation sessions.

My next question explored adoption and privacy behaviors over the course of the semester.

Freshman Facebook Participation, Fall Semester 2005

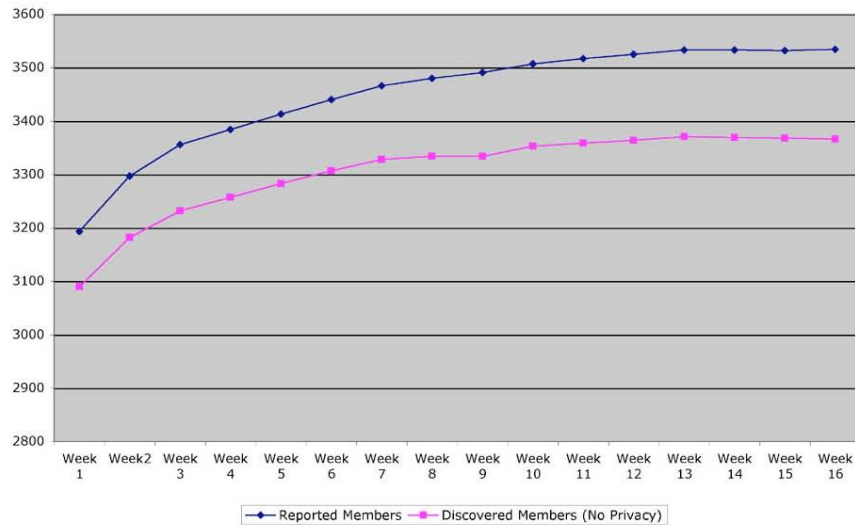


Figure 2. Freshman Facebook Participation, Fall Semester 2005.

On the first day of school, 3193 freshmen possessed a Facebook account. This was over 85% of the entire freshman class; use grew to over 94% by the end of the semester. The blue line represents total accounts, and the pink line reflects total accounts that are not private, meaning anyone with a [X University] Facebook account can view that profile. As the semester passed, students protecting their profile grew from 3.2% to 4.75%.

Finally, I wished to explore the growth of the aggregate social network of freshmen over the fall semester.

The Social Network Expansion of Freshmen

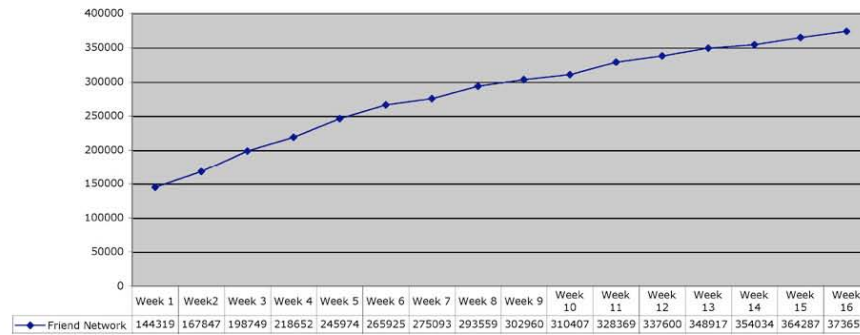


Figure 3. Social Network Expansion of freshmen

While the actual number of nodes (the freshmen) in the network did not grow substantially over the course of the semester, the number of edges (friendship connections) in the network did expand remarkably. As the freshmen made friends over the course of the semester, their social network size grew from 144,319 to 373,651 connections. The average number of friends a freshman on the Facebook had on day one was 46, and at the end of the semester, he or she had 111 friends. This might give us a picture of how many “connections” a freshman might make the first semester of college: 65.

In the poster, I will expand substantially on these preliminary findings, exploring the dynamics of profile structure over the course of the semester. Furthermore, I will explore significant correlations between profile construction and friendship behaviors, furthering the work of Lampe, Ellison and Steinfeld (2007). While the findings will be primarily descriptive, the unique nature of this dataset and the time in which it was captured promise valuable insight into social network site behavior.

Citations

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In P. Golle & G. Danezis (Eds.), *Proceedings of 6th Workshop on Privacy Enhancing Technologies* (pp. 36-58). Cambridge, UK: Robinson College.

boyd, d. (2006). Friends, Friendsters, and MySpace Top 8: Writing community into being on social network sites. *First Monday*, 11 (12). Retrieved October 28, 2007 from http://www.firstmonday.org/issues/issue11_12/boyd/.

Hays, R. & Oxley, D. (1986) Social Network Development and Functioning During a Life Transition. *Journal of Personality and Social Psychology*, 50 (2), 305-313.

Lampe, C., Ellison, N., & Steinfield, C., (2006). A Face(book) in the crowd: Social searching vs. social browsing. *Proceedings of CSCW-2006* (pp. 167-170). New York: ACM Press.

Lampe, C., Ellison, N., & Steinfield, C. (2007). A familiar Face(book): Profile elements as signals in an online social network. *Proceedings of Conference on Human Factors in Computing Systems* (pp. 435-444). New York: ACM Press.

POSTER SUBMISSION 18:

Healthcare Informatics: Supporting Collaborative Sensemaking in the Emergency Department

AUTHOR(S):

Paul, Sharoda A.

ORGANIZATION(S):

Pennsylvania State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Poster for iConference 2008

Oct 29, 2007

Healthcare Informatics: Supporting Collaborative Sensemaking in the Emergency Department

It is a busy day in the emergency room with the monitors beeping, the alarms blaring intermittently, the phones ringing, and the overhead paging system crackling at intervals with important messages. The nurses are running back and forth between the rooms, weaving around the patients on the hallway beds. The doctors are talking in groups while referring to the large overhead displays showing the status of patients. The charge nurse is poring over her computer screen, trying to find beds for waiting patients, while talking on the phone.

A paramedic walks up to one of the registration associates and asks, "Do you know what the trauma number is for the patient that just came in?"

The registration associate quickly glances through the list of patients on her computer screen and points to a record, "Is that the one? Came in about 30 minutes ago?"

The paramedic looks at the record and shakes his head, "No this just came in a few minutes ago. 17-year-old motor vehicle accident."

The registration associate looks through the list again, and shakes her head, "I don't see any other trauma patient in here."

The paramedic looks at her, frustrated, "I just brought this patient in. Now I can't find him." He walks away, looking for his patient.

– Field-notes from the emergency department at the Hershey Medical Center

PROBLEM MOTIVATION

Emergency Department Overcrowding

In the face of rising costs and growing demand, the healthcare industry is facing immense challenges today with respect to quality and efficiency of care delivery. This is reflected in the dismal state of the emergency departments of hospitals across the country. The emergency department (ED) provides care for acutely ill and injured patients 24 hours a day, 7 days a week. The ED also provides care to under-served populations who have no other options for medical care [1] and has thus been designated the 'safety net' of the healthcare system [2].

Over the past three decades, EDs have been facing an overcrowding problem and this has now reached crisis proportions [1, 3]. Between 1993 and 2003, ED visits went up by 23.6 million, while at the same time 425 EDs closed and hospital beds declined by 198,000 [5]. ED overcrowding manifests itself in many ways: "boarding" of patients in the ED, ambulance diversions, patients being treated in hallways, and patients leaving without treatment [1, 6]. Overcrowding leads to long wait times, increased medical errors, poor patient outcomes, high levels of stress among ED staff and decreased capacity of the nation's EDs to respond to mass casualty incidents [5].

In the healthcare field, research aimed at addressing the problem of ED overcrowding has fallen primarily into three categories – descriptive, predictive and intervention-oriented. Descriptive studies have focused on defining overcrowding [7], examining its causes and effects [8] and developing measures to quantify it [9, 10]. Predictive studies have focused on developing early warning systems for impending overcrowding episodes [11]. A third stream of research has focused on interventions to optimize available resources and processes, such as monitoring patient length of stay, and re-designing

Poster for iConference 2008

Oct 29, 2007

processes and patient flows [12, 13]. Information and communication tools (ICTs) can play an important role in this line of research.

The Role of information and communication technologies

In the ED, ICTs have the potential to increase quality and efficiency of care by enabling information sharing, communication, collaboration, and coordination between care providers [5]. Tools such as the electronic medical record, the computerized provider order entry system, and the electronic bed tracking system can help ensure smooth flow of information and patients in the ED and between the ED and the rest of the hospital. Given the push in healthcare towards deployment of technologies to streamline processes, reduce medical errors, and provide better collaboration for providers, ICTs can help alleviate overcrowding of hospital emergency departments.

PRELIMINARY EXPLORATION

Research Question

I started exploring the role of technology in ED overcrowding with the following research question:

RQ1: "How can information and communication technology used in emergency departments be leveraged to alleviate levels of overcrowding"

Methods

I conducted a field study using ethnographic techniques including observations and interviews. The field research site was chosen to be the Penn State Hershey Medical Center (HMC). HMC is a 500-bed teaching hospital with nearly 50,000 ED visits a year. The ED at HMC is a level-1 trauma center and is serviced by a critical care transport service consisting of medically-equipped helicopters, a ground service comprising nine ambulances and a pediatric ambulance. HMC staff use an integrated electronic medical record that interfaces with lab, radiology, and pharmacy information systems and the computerized provider order entry system; they also use communication tools like pagers, radio, phones etc.

I conducted preliminary fieldwork from Jan – May 2007 aimed at understanding the nature of work in the ED. I focused on the roles and responsibilities of care providers, the patient flows, the information flows, the resource flows and the communication flows.

Preliminary results: Importance of collaborative sensemaking

At the end of my preliminary fieldwork I analyzed the data using a grounded theory approach [14] to identify interesting themes. A recurrent theme was the need for providers to understand dynamic situations in collaboration with other providers and using various paper and digital tools. This brought to the forefront the issue of *collaborative sensemaking* in the emergency department.

Sensemaking involves finding structure in a seemingly unstructured situation [15] and is an integral part of the work of care teams in the ED. *Collaborative sensemaking* occurs when multiple actors with different thoughts about the world engage in the process of understanding 'messy' data or information [16]. Providing care to patients in the ED

Poster for iConference 2008

Oct 29, 2007

requires collaboration between physicians, nurses, technicians, special consults and administrative staff. Given the variety of patients coming to the ED, the time-criticality of action, the unavailability of information and the need for collaboration among providers with different backgrounds, collaborative sensemaking is an important aspect of work in the ED. An instance of this is shown by the vignette presented at the beginning of this paper.

In their work, care providers use information tools like paper patient charts, whiteboard, electronic medical records, computerized provider order entry system and communication tools like phones, and pagers. I found that different providers leverage these tools in different ways to support their sensemaking activities. At the end of my preliminary fieldwork, I re-focused my research question as follows:

RQ2: "How can information and communication tools used in the emergency department be leveraged to support collaborative sensemaking among healthcare providers?"

Sensemaking research

The process of how groups make sense collaboratively is not clearly understood. At the organizational level, Weick [17] characterized sensemaking as having seven characteristics – it is grounded in identity construction, retrospective, focused on and by extracted cues, enactive of sensible environments, social, ongoing, and driven by plausibility rather than accuracy. Russell et al. [18] define sensemaking as cyclic processes of searching for external representations and encoding information into these representations to reduce the cost of tasks to be performed. While some tools [19-21] have been developed to enhance sensemaking, there is little understanding of and technological support for the process of sensemaking in groups, specifically in time-critical situations such as emergency response.

PROPOSED DISSERTATION WORK

In my dissertation, I propose to a)examine the process of collaborative sensemaking, b)examine if and how currently used technologies and interfaces support collaborative sensemaking, and c)identify requirements for developing technologies to support collaborative sensemaking activities. Given these goals, I am interested in answering the following questions in the context of the emergency department through ethnographic observations, interviews, shadowing, and artifact collection.

- RQ 2.1 What are the aids and approaches used by groups for sensemaking?
- RQ 2.2 How do existing ICTs used in the ED support collaborative sensemaking?
- RQ 2.3 How can ICTs be designed to support collaborative sensemaking in time-critical work?

CONTRIBUTION

The proposed research has the potential to make significant contributions in three fields: medical and healthcare informatics, computer-supported cooperative work (CSCW), and human-computer interaction (HCI). The contribution to the healthcare informatics community will be in determining how ICTs can improve the quality and efficiency of

Poster for iConference 2008

Oct 29, 2007

care in emergency departments. The contribution to the CSCW community will be in helping understand the processes and aids used in collaborative sensemaking, which can feed into design requirements for collaborative systems. Finally, the contribution to the HCI community will be in finding how information can be visualized and represented in interfaces of systems used for sensemaking.

CONCLUSION

In this paper I've highlighted my dissertation work in which I propose to tie research on computer-supported cooperative work (CSCW) to the real-world problem of overcrowding in hospital emergency departments. For long CSCW has studied how groups use technology in real-world settings and how technology can be better designed to support groups working in these work settings. In my dissertation I propose to examine a) the process of collaborative sensemaking in groups, b) how ICTs are used currently to support collaborative sensemaking, and c) how new tools can be designed to support collaborative sensemaking.

Poster for iConference 2008

Oct 29, 2007

REFERENCES

1. Trzeciak, S. and E.P. Rivers, *Emergency department overcrowding in the United States: an emerging threat to patient safety and public health*. *Emergency medicine journal*, 2003. 20: p. 402-405.
2. IOM Committee on the Changing Market, M.C., and the Future Viability of Safety Net Providers., *America's Health Care Safety Net: Intact but Endangered*, ed. M. Lewin and S. Altman. 2000, Washington, D.C., USA: National Academies Press.
3. Richardson, L. and U. Hwang, *America's Health Care Safety Net: intact or unraveling*. *Academic Emergency Medicine*, 2001. 8(11): p. 1056-63.
4. Derlet, R.W., *Overcrowding in emergency departments: increased demand and decreased capacity*. *Annals of Emergency Medicine*, 2002. 39(4): p. 430-432.
5. IOM, *Hospital-based emergency care: At the breaking point*. Vol. Pre-publication copy. 2006, Washington, D.C.: National Academies Press.
6. Fatovich, D.M. and R.L. Hirsch, *Entry overload, emergency department overcrowding, and ambulance bypass*. *Emergency Medicine Journal*, 2003. 20: p. 406-409.
7. Hwang, U. and J. Concato, *Care in the Emergency Department: How Crowded is Overcrowded?* *Academic Emergency Medicine*, 2004. 11(10): p. 1097-1101.
8. Derlet, R.W. and J.R. Richards, *Overcrowding in the Nation's Emergency Departments: Complex Causes and Disturbing Effects*. *Annals of Emergency Medicine*, 2000. 35(1): p. 63-68.
9. Bernstein, S.L., et al., *Development and Validation of a New Index to Measure Emergency Department Crowding*. *Academic Emergency Medicine*, 2003. 10(9): p. 938-942.
10. Weiss, S.J., et al., *Estimating the Degree of Emergency Department Overcrowding in Academic Medical Centers: Results of the National ED Overcrowding Study (NEDOCS)*. *Academic Emergency Medicine*, 2004. 11(1).
11. Hoot, N. and D. Aronsky, *An Early Warning System for Overcrowding in the Emergency Department*, in *AMIA Annual Fall Symposium*. 2006, American Medical Informatics Association: Washington, D.C. p. 339-343.
12. Schneider, S., et al., *Rochester, New York: A Decade of Emergency Department Overcrowding*. *Academic Emergency Medicine*, 2001. 8(11): p. 1044-1050.
13. Spaite, D.W., et al., *Rapid process redesign in a university-based emergency department: decreasing waiting time intervals and improving patient satisfaction*. *Annals of Emergency Medicine*, 2002. 39(2): p. 168-177.
14. Strauss, A. and J. Corbin, *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. 1990, Newbury Park, CA.: Sage Publications.
15. Furnas, G.W. and D.M. Russell. *Making sense of sensemaking*. in *CHI '05 extended abstracts on Human factors in computing systems*. 2005. Portland, OR, USA: ACM Press.
16. Ntuen, C.A., P. Munya, and M. Trevino. *An approach to collaborative sensemaking process*. in *Proceedings of the 11th International Command and Control Research and Technology Symposium*. 2006. Cambridge, UK.
17. Weick, K.E., *Sensemaking in Organizations*. *Foundations for Organizational Science*, ed. D. Whetten. 1995, Thousand Oaks, CA: Sage Publications, Inc. 231.

Poster for iConference 2008

Oct 29, 2007

18. Russell, D.M., et al. *The cost structure of sensemaking*. in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1993. Amsterdam, The Netherlands: ACM Press.
19. Qu, Y. *A Sensemaking Supporting Information Gathering System*. in *CHI '03 extended abstracts on Human factors in computing systems*. 2003. Ft. Lauderdale, Florida, USA: ACM Press.
20. Keel, P., *EWall: A visual analytics environment for collaborative sensemaking*. *Information Visualization*, 2007. 6: p. 48-63.
21. Billman, D. and E.A. Bier. *Medical Sensemaking with Entity Workspace*. in *Proceedings of the SIGCHI conference on human factors in computing systems*. 2007. San Jose, CA: ACM Press.

- 6 -

POSTER SUBMISSION 19:

Conversation Repository for Participatory Librarianship

AUTHOR(S):
Inoue, Keisuke

,

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

A Conversation Repository for Participatory Librarianship

Abstract

The goal of this study is to model conversation data from digital reference services and reconstruct the processes of the reference interviews as a conversation repository. By doing so, I am hoping that the study will provide an infrastructure for applications such as an interactive information retrieval system, a question answering system, or a recommender system for information intermediaries. While the current study focuses on the digital reference conversation, observations may be generalizable, to a certain extent, to different online conversations such as blogs, question answering sites, social networking sites, or product review systems, and thus contribute to a wider range of knowledge sharing and knowledge creation through information technologies.

This work is a part of the Participatory Librarianship research initiative, proposed by Lankes et al. (2006), which seeks an implementation of library as a facilitator of technology-enhanced human communication and knowledge creation, based on Gordon Pask's Conversation Theory. (Pask, 1975, 1976)

Motivation

While fully automated Web search engines have become the primary tool for information searching of the majority of Web users, human-in-the-loop systems such as product review systems (e.g. Amazon), community-based QA systems (e.g. Yahoo! Answer), and online chat services with information experts (e.g. help desk and virtual reference services) have also been increasing in popularity in recent years. (Resnick et al., 2000; Madden and Fox, 2006) These systems, which became available largely through the development of Web 2.0 technologies, provide an environment for information seeking dialogues and social networking opportunities by employing easy-to-use interfaces for interactive communications among the users in a public domain or a closed shared domain. Since the questions and answers are exchanged in natural language texts, as opposed to query terms and hyperlinks as in the case of Web search engines, the messages exchanged in these systems are richer and more complex in their information. Moreover, since the communications are made online, the transcripts of the conversations can be stored for future reference. However, as rich and complex as the natural language conversations are, searching through an archive of conversations is a difficult task, because of its linguistic structure being different from other kind of text data such as books or documents, where information retrieval has a long tradition of research.

For example, in an information-seeking conversation, even the description of the information sought in the conversation may not always be available. It is known that one of the challenges of a reference interview is to find out what the user is asking for. (Taylor, 1968; Belkin et al., 1982). According to the ASK Hypothesis (Belkin et al., 1982), users are often unable to identify their information need at the beginning of an interview, and the information requirement becomes overt only through a successful information-seeking dialogue, which consists of iterative interactions between the user and the information intermediary with a certain structure. Early studies in information behavior (Oddy, 1977; Brooks and Belkin, 1983; Daniels et al., 1985; Brooks et al., 1986; Belkin et al., 1987) confirmed this hypothesis, by revealing the discourse structure of various reference interviews. For example, Daniels et al. (1985) analyzed face-to-face interviews at the reference desk in various research institutes at London University and identified the *problem structure* of reference conversation, which organized the goals of utterances in the interviews hierarchically. However, the focus of the majority of IR system research has been experimental studies based

on the “single-shot” search and evaluation, and a few studies have incorporated the notion of iterative interaction. Relevance feedback is an active field of study in IR, which employed iterative feedbacks from the user, but as Bates (1989) pointed out, the method assumes that the user’s information need never changes throughout the search process and thus sits well in the classic model of IR systems.¹

On the other hand, the library has long been a place for information-seeking interactions, most explicitly through reference services, but also through other functions or events such as building collections based on community needs or organizing book groups and speaker series. Searching books by browsing through book shelves is also a form of interactive information searching: the user can look for books in different directions based on the evolving information need: from a broad topic to a narrower one or old books to newer ones. Collecting information piece by piece, known as *berypicking* (Bates, 1989), which is an important capacity of information-seeking interactions, is also available in this mode of information searching.

The Participatory Librarianship research initiative, proposed by Lankes et al. (2006), seeks an implementation of library as a facilitator of technology-enhanced information interactions. It is motivated by Gordon Pask’s conversation theory (Pask, 1975, 1976), which posits that knowledge is created through conversations. A conversation repository is an infrastructure to facilitate such conversations by incorporating theories from information behavior study, (ASK hypothesis, *berypicking* model), cybernetics (conversation theory), and discourse linguistics (Grosz and Sidner model) into the emerging participatory information systems that has been enabled by the Web 2.0 technologies.

Current Approach: Discourse Analysis for Conversation Repository

While the intentions of these theories which support knowledge creation and information discovery through iterative interactions resonate with the spirit of Participatory Librarianship, the technical aspects of them have not developed enough to be compatible with the emerging participatory information technologies. However, we believe that the methods of discourse analysis can complement and modernize the earlier work by revealing the structure of information-seeking conversations and making it available as an input to information systems to enhance information retrieval processes.

Discourse analysis is an approach for studying instances of language beyond the sentential level, such as utterances in conversation. Linguist, Harris (1952) defined the term as a formal method for linguistically analyzing “connected speech (or writing)”, but the approach has been applied to various fields in social science, where the focus is on the relations between the language and its effects to the society. For understanding of conversation in the information seeking context, Grosz and Sidner (1986) developed a framework that analyzes the discourse structure by three distinctive (but interrelated) components: linguistic structure, intentional structure, and attentional state. Their approach has been directly or indirectly applied in numerous studies. (Daniels et al., 1985; Carberry and Lambert, 1999; Carberry et al., 2006; McTear, 2002)

The study will analyze the discourse structure of the virtual reference interviews based on their framework. Specifically, the study will seek answers for the following questions:

- What are the linguistic characteristics of digital reference interviews?

¹The HARD track for TREC (Allan, 2005) and studies in Cognitive Information Retrieval (Spink and Cole, 2005) are some of the exceptions.

- What are the intentions of utterances in digital reference interviews and how are they related to the linguistic characteristics?
- How do the attentions change in digital reference interviews and how are they related to the linguistics characteristics?
- How is the discourse structure of digital reference interviews different from other types of conversation?

In addition, the study will examine the following questions in order to contribute to implementing conversation-based information systems.

- What are the machine learning techniques that suit for learning the discourse structure of information seeking dialogues?
- How can the automatic detection of discourse structures be utilized in the information seeking process?

Data

Our current data, provided by the Online Computer Library Center (OCLC), is a log of 450 virtual reference service interviews, consisting of 8066 lines of messages exchanged by various users and librarians from all over the United States. The reference interviews took place in the form of on-line chat sessions with a co-Web-browsing capability.

Conclusive Remarks

The combination of conversation theory, information behavior study, discourse analysis, and emerging information technologies have enabled us to design our research in the following ways:

- Pask's conversation theory provides a conceptual framework to build an information repository that emphasizes knowledge creation through conversations.
- The early studies of information behavior provides a theoretical motivation for incorporating the notion of conversation into information retrieval technologies, enabled by the emerging participatory information systems.
- Discourse analysis methods can be used to reveal the structure of conversation.
- These discourse models will inform the development of conversation repository, an infrastructure for conversation-based information systems.

While the study currently focuses on the digital reference conversation, the observed discourse structure may be generalizable, to a certain extent, to different kinds of conversation. Especially, the conversations through the participatory information technologies, such as community-based question answering systems, blogs, social networking sites, or product review systems, may provide a wide range of complex and detailed daily-life information that has not been easily searchable. Therefore, incorporating the observations from this study to these conversations will contribute to a wider range of knowledge sharing and creation through information technologies.

References

- Allan, J. (2005). Hard track overview in trec 2005 high accuracy retrieval from documents. In Voorhees, E. M. and Buckland, L. P., editors, *Proceedings of the Fourteenth Text REtrieval Conference Proceedings*.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5):407–424.
- Belkin, N. J., Borgman, C. L., Brooks, H. M., Bylander, T., and Croft, W. B. (1987). Distributed expert-based information systems: an interdisciplinary approach. *Inf. Process. Manage.*, 23(5):395–409.
- Belkin, N. J., Oddy, R. N., and Brooks, H. M. (1982). Ask for information retrieval. part i: Background and theory. *Journal of Documentation*, 38(2):61–71.
- Brooks, H. M. and Belkin, N. J. (1983). Using discourse analysis for the design of information retrieval interaction mechanisms. In *SIGIR '83: Proceedings of the 6th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 31–47, New York, NY, USA. ACM Press.
- Brooks, H. M., Daniels, P. J., and Belkin, N. J. (1986). Research on information interaction and intelligent information provision mechanisms. *J. Inf. Sci.*, 12(1-2):37–44.
- Carberry, S., Elzer, S., and Demir, S. (2006). Information graphics: an untapped resource for digital libraries. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 581–588, New York, NY, USA. ACM Press.
- Carberry, S. and Lambert, L. (1999). A process model for recognizing communicative acts and modeling negotiation subdialogues. *Comput. Linguist.*, 25(1):1–53.
- Daniels, P. J., Brooks, H. M., and Belkin, N. J. (1985). Using problem structures for driving human-computer dialogues. In *In RIAO 85, Acts of the Conference: Recherche d'Informations Assistee par Ordinateur*, pages 645–660, Grenoble. I.M.A.G.
- Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Harris, Z. S. (1952). Discourse analysis. *Language*, 28(1):1–30.
- Lankes, R. D., Silverstein, J., and Nicholson, S. (2006). *Participatory Networks: The Library as Conversation*. Information Institute of Syracuse.
- Madden, M. and Fox, S. (2006). Riding the waves of “web 2.0”: More than a buzzword, but still not easily defined.
- McTear, M. F. (2002). Spoken dialogue technology: enabling the conversational user interface. *ACM Comput. Surv.*, 34(1):90–169.
- Oddy, R. N. (1977). Information retrieval through man-machine dialogue. *Journal of Documentation*, 33(1):1–14.

- Pask, G. (1975). *Conversation, cognition and learning*. Elsevier, New York.
- Pask, G. (1976). *Conversation Theory Applications in Education and Epistemology*. Elsevier.
- Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12):45–48.
- Spink, A. and Cole, C. (2005). *New Directions in Cognitive Information Retrieval*. Springer.
- Taylor, R. T. (1968). Question-negotiation and information seeking in libraries. *College and Research Libraries*, 29:178–194.

POSTER SUBMISSION 20:

A preliminary consumer health information-seeking (CHI-seeking) behavior model of physicians who treat elderly depressed patients (results from the pilot study of a dissertation proposal)

AUTHOR(S):

Dorsey, Mary Jo

ORGANIZATION(S):

University of Pittsburgh, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Objective:

Despite useful studies that report the information-seeking behaviors of physicians who use the Internet, or physicians' clinical information-searching skills (such as evidence-based medicine), little literature highlights physicians' consumer health information-seeking behavior. Where does a physician get information to give to the patient or to a caregiver? In a rapidly growing market of consumer health information, what does a physician offer patients today? This poster describes pilot research, methodology and expectations of a study designed to investigate information-seeking behaviors of primary care physicians who treat elderly depressed patients. A main objective is to reveal a correlation between what information is sought versus what is provided to a patient or caregiver. This study identifies a specific population, senior health consumers, and investigates the steps their front-line outpatient caregivers (primary care or family practice physicians) take in order to make sure their patients have the information they need to understand their illnesses and to take part in a shared decision-making process with regard to their treatment. The study attempts to uncover the consumer information-seeking habits of primary care physicians who see large populations of elderly depressed patients. Primary research questions include:

1. How are these physicians searching for consumer health information?
2. How are physicians responding to health information requests from their patients or caregivers?
3. Can a consumer health information-seeking model that is significant for the medical and LIS professions be constructed?

Setting/subjects:

Physicians who treat seniors with mood disorders were selected for two reasons. First, senior citizens (those aged 65+) are a major demographic in today's society. Second, access to populations of both primary care physicians and family practice physicians who are affiliated with the world-class center of care for elderly depressed patients and their caregivers is available at the University of Pittsburgh. Three specific procedures, structured interviews, environmental scans, and a self-evaluative confidence scale, will take place with physicians in large ambulatory primary care practices in urban Pittsburgh, Pennsylvania, who take care of large numbers of geriatric patients. In the larger study, twenty-five physicians will be interviewed from Community Medicine, Inc. and Metro Family Practice, two practices associated with the Advanced Center for Intervention and Services Research for Late-Life Mood Disorders (ACISR) at the University of Pittsburgh, as well physicians from the Department of Family Medicine at the University of Pittsburgh School Of Medicine. In the pilot study, three family practice physicians from the Department of Family Medicine at the University of Pittsburgh School of Medicine were interviewed and administered a confidence level survey. The interviewer also performed an environmental scan of the clinical setting according to a predetermined checklist of items to note during the interview. University of Pittsburgh Institutional Review Board approval was secured prior to the investigation.

Methodology:

Using a grounded theory approach similar to that used in Ellis' information-seeking behavior model research of 1989, 1993 and 1997, [1, 2, 3, 4] data was collected through personal interviews with Pittsburgh-area primary care physicians who see large numbers of elderly patients in their practices, together with environmental scans of their practice spaces and their responses to a standardized questionnaire on their confidence in using consumer health information with their patients. Glaser & Strauss' "constant comparative method" [5] will lead to a generation of theory about the physicians' common information-seeking behaviors from the empirically collected data. Data will be gathered in three phases. (1) Semi-structured interviews, (2) Environmental scans, (3) Self-evaluative confidence scale.

Results:

Use of a naturalistic inquiry method inspired by Ellis [4] will be used to synthesize the use of grounded theory approaches and related methodological issues for the specific purpose of creating a consumer health information-seeking behavior model of primary care physicians who treat elderly patients, and, in turn, will be used to educate clinicians and develop further information-seeking skills of medical students, interns and residents who are training

to become better information providers for their patients. Through environmental scans of their offices and interviews with these physicians who are affiliated with the ACISR, outcomes of the pilot include a preliminary model of (1) common behaviors among this relatively homogenous group of physicians, and (2) how these physicians respond to their elderly patients and their caregivers with respect to the provision of consumer health information. The preliminary model indicates three stages of information-seeking. (1) The pre-patient visit includes tasks such as listening to or reading current medical news that patients might be hearing, perusing websites that their patients (or their caregivers) may be familiar with, looking for education materials on a case-based nature. (2) Information-seeking activities that occur during the patient visit primarily include reaching for the computer in the clinical office with the patient present and searching for a website or searching UpToDate or MDConsult for information to show the patient. Bulleted text and easy to read websites are preferred. During this stage, the physician gives the patient and caregiver website addresses written on a prescription pad to take home for further use. (3) Information-seeking behaviors which occur following a patient visit include website follow-up or notes to self in a palm or hand-held device to gather information at a later time or request a search from the library or an administrative assistant. Scans of the physicians' clinical environments revealed electronic equipment such as computers, hand-held devices and electronic resources such as Internet access and use of MDConsult, UpToDate or other specific consumer health resources were of particular interest. More obvious were textbooks, pamphlets, journals and wall hangings which were noted, in detail, in the environmental scan notes. Results of the confidence scale administered to the physicians during the pilot study showed confidence levels that ranged from very confident need to improve; to confident but should improve; to not confident but should improve. The majority of questions resulted in confident but should improve, thus initially indicating that although physicians are confident in providing information to their patients, they still feel that there is more to know about consumer health information.

Discussion/conclusion:

The preliminary model. The preliminary consumer health information-seeking model of physicians interviewed during the pilot stage of this study begins by dividing the information-seeking into three stages described as a PDA, where the P is for Prior (pre-patient visit, anticipating a need), the D is for During (while patient is in office, direct question and answer session), and the A is for After (look at websites suggested by patients after patient leaves). Practical uses for a model of physicians' information-seeking behaviors include awareness and education. Awareness of trends of information habits can lead to awareness of one's own information needs and pursuits. Medical school educators instruct physicians in training, LIS faculty educate graduate library and information science students who, in turn as medical librarians, teach information seeking and retrieval methods to physicians. A synthesized behavior model can enhance this educational process. Finally, a specific goal of this particular group of physicians (in Pittsburgh) is that if an understanding of the CHI information-seeking behaviors of this group of physicians is successful, then the ACISR can also create their own consumer health information materials based more precisely on their patients' and caregivers' information needs.

References

1. Ellis, D., Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *Library Quarterly*, 1993. 63(4): p. 469-486.
2. Ellis, D., A Behavioral Model for Information Retrieval System Design. *J Information Science*, 1989. 15(4/5): p. 237-247.
3. Ellis, D., D. Cox, and K. Hall, A comparison of the information seeking patterns of researchers in the physical and social sciences. *J Documentation*, 1993. 49(4): p. 356-369.
4. Ellis, D., Modeling the information seeking patterns of engineers and research scientists in an industrial environment. *J Documentation*, 1997. 53(4): p. 384-403.
5. Glaser, B. and A. Strauss, The discovery of grounded theory; strategies for qualitative research. *Observations*. 1967, Chicago: Aldine Pub. Co. 271p.

POSTER SUBMISSION 21:

The impact of documentation on secondary data use

AUTHOR(S):

Niu, Jinfang

ORGANIZATION(S):

University of Michigan, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Introduction

Organizing, managing information, and providing information services to users have been long traditions in information schools. Numerous researches have been done on the organization and use of books, journals, images, even audio and video resources. Social science raw data, on which many publications and scientific findings are based, are important information resources. But they have received inadequate attention from information schools in the past. This is partly because the secondary use of raw data was not a common practice in many academic fields. Facilitating data sharing has been a growing concern in recent years. In the United States, a law has been passed by the Congress to mandate the Office of Management and Budget to amend Circular A-110¹ to extend the Freedom of Information Act (FOIA) to "require Federal awarding agencies to ensure that all data produced under an award will be made available to the public under the FOIA" (<http://www.whitehouse.gov/OMB/fedreg/a-110rev.html>). More and more funding agencies require grantees to share their research data to the public. We can expect that in the near future, more and more social science raw data will be available for public access and use. Thus, there needs to be more research about the organization, management and use of social science data.

Secondary data use is defined this way: a user uses a data set, and the user is not involved in the production process of that data set. Three basic entities are involved in secondary data use: the data producers, data users, the data itself and associated documentation. Sometimes there is an intermediary between data producers and users.

Intermediaries, such as data archives or data libraries, process the data to improve the quality of data and documentation, and disseminate data to users.

Documentation is metadata of social science data. Similar to MARC records that help users to search and judge the relevance of books and journals, Dublin Core records help users to search and determine the relevance of Web resources. Documentation provides information necessary to search and judge the relevance of data, more importantly, they help users understand and use the data. The Interuniversity Consortium for Political and Social Research (ICPSR) identifies the following as necessary elements of good documentation: cataloging information, such as title, principal investigator, data producer, place and date of production, funding agency; description of how the data were collected and the data sources used; full description of sampling design, frame, and methods as well as sampling error; full variable and value labels, full details of all coding classifications; question text, full description of recoded and derived variables, frequencies of variables, fully documented weights with information on conditions under which they should be used; and details on file types and linkages among files. If available, data collection instruments and related bibliographies should also be provided.

Research Questions

There is a lack of empirical studies on the secondary use of social science data. Existing literatures about social science data sharing were mostly written in the 80's and 90's. More recent studies on the subject seem to be lacking. Most of the available literature discusses the need for data sharing, the costs, benefits and risks of data sharing, obstacles to share data, the legal environment of sharing social science data, what intermediaries do to help data sharing, etc. Some papers explore use experiences, but these mainly consist of individual opinions and arguments based on personal stories or anecdotal evidences, or summaries of discussions at conferences. While most literature mention that documentation is often insufficient for secondary use, they did not explain how the insufficiency affects secondary data use, or how users deal with insufficiency.

This study shall fill this gap by doing a systematic and empirical study on the secondary use of social science data, focusing on the impact of insufficient documentation. It breaks use experiences into three aspects. First, users choose which data set to use. Second, users get the data, try to use it, and then decide to give up or continue to use it. Third, how difficult users feel about their using experiences. The aim is to examine the effects of documentations on each of those three aspects. The research questions are framed as follows:

1. How do users choose which data to use?
2. What are the reasons why users give up using or continue to use a data set?
3. What factors make documentation more or less sufficient for use?
4. What factors make data use experiences more or less difficult?

¹ Circular A-110 is a policy that implements FIOA.

Methodology

Three steps are involved in answering these questions. First, from existing studies on knowledge reuse, knowledge transfer and the secondary use of non-social science data, a list of factors for question 3 and 4 has been identified. Second, qualitative interviews shall be conducted. Based on the interviews, answers to research question 1 and 2 shall be explored. In addition, the preliminary list of factors shall be refined by adding and deleting factors, breaking one factor into several sub-factors, or adjusting the factors to make them applicable for the secondary use of social science data. For the factors that will be analyzed quantitatively, quantifiable measures shall be established. In the end, a survey will be conducted

Respondents of study shall be the end users of secondary social science data. For the interviews, the convenient sample and snowball sampling methods shall be employed. A preliminary step is to contact known secondary data users. Further respondents shall be contacted based on the names recommended by the initial group. Respondents to the survey shall be drawn from the user pool of a data archive.

For the first two research questions, the units of analysis are the individual data users. For the third and fourth research questions, the units of analysis are the single data sets in individual use cases. For each interviewees or survey respondents, general questions about their secondary data use experiences shall be asked to generate the answers to research question 1 and 2. Then, they will be asked to talk about their recent experiences in using a particular data set. Through this procedure, data on two kinds of units of analysis can be generated through one study.

The first two research questions will be answered qualitatively. Question 3 and 4 will be answered both qualitatively and quantitatively. For the quantitative analysis of question 3, the dependant variable is the sufficiency of documentations. Below is the list of independent variables identified so far:

- Source of data: where the data is obtained from;
- Type of data: is the data survey/interview data, experimental data, longitudinal data etc.
- Producers of data: is the data produced by individual researchers, private organizations, or government agencies.
- The time of data production: when the data was produced.
- Knowledge distance: how much are the users familiar with the data, the topic of the data, the data collection and analysis methods.
- Collaboration: how many other people are the users collaborated in using the data.

For research question 4, the dependant variable is the level of difficulty, which is users' perception of how difficult it was using the data. The independent variables are: the sufficiency of documentations, all the independent variables for question 3, plus two additional variables, the popularity and complexity of the data. Popularity is measured by the number of publications based on the data. The complexity of the data set is measured by the number of sub data sets, variables and observations in the data set, the actual number of variables and observations used by the users.

Significance of the Study

This is an attempt to provide the first systematic study on the secondary use of social science data. The empirical findings will benefit two parties. First, it tells secondary users what kind of data are better documented and easier to use, it also tells them how they should prepare themselves to make their data using experience easier. Second, it will inform data libraries or data archives what kind of data need more processing and what they can do to help users. This study categorizes social science data and its users, identifies factors affecting using experiences, and describes relationships among factors. In this sense, it is a step towards a theory of secondary data use. Moreover, I applied knowledge management theories into secondary data use. Findings of this study will contribute to the theories of knowledge reuse and knowledge management.

POSTER SUBMISSION 22:

Cyber Java Monopoly: Game-based approach of collaborative programming language learning

AUTHOR(S):
HSIAO, I-Han

LIN, Yi-Ling

ORGANIZATION(S):

University of Pittsburgh, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The Impacts of Gender and Initial Social Activities on Trust and Performance in CMC

1 Introduction and Research Questions

In recent years HCI researchers have made significant progress in understanding how different communication media influence people's trust perceptions and task performance in virtual environments [1, 5]. In this research little attention has been paid to the effects of gender differences in communication in virtual environments. However, there is good reason to investigate the role of gender, given research in the discipline of communication studies that has shown gender effects in face-to-face and virtual communications. In communication studies, gender analyses have focused on patterns of communication involving status, power, and influence [3]. By contrast, this research aims to bring the gender factor into computer-mediated communication (CMC) from an HCI perspective, responding to central concerns of efficiency, effectiveness, and user perceptions.

The primary objective of this research is to explore gender differences in synchronous computer-mediated communication with and without initial social activities. In particular, I ask whether initial social activities affect trust development and performance of male and female pairs in a social dilemma game.

To this end, the research questions for this study are:

- 1) How does gender influence people's trust development over multi-trials of a competitive task via the IM System?
- 2) How effectively does pre-task activity, i.e. social chat, help different gender pairings achieve higher levels of trust and better performance over multi-trials of a competitive task?
- 3) Is there a larger benefit of pre-task activity for males, who tend to be less trusting [4], than for females in doing a task which involves conflicts of interest? If the answer is YES, do males achieve the same level of trust as females?
- 4) Qualitatively, how do different gender pairings use language to communicate via the IM System and how does their language affect trust?

In order to get answers for these questions, I measure three dependent variables: participants' post-task trust, participants' performance and behaviors during the task and the pairs' communication processes.

2 Methodology and Primary Results

A 2x2 between subjects design was used: gender pairing (male/male vs. female/female), and pre-task interaction (interaction vs. no interaction). All participants used the AOL IM System to do the task. The pairs were strangers and the gender of the partners was revealed by the experimenter immediately before pairs performed their task.

Participants

A total of 124 participants have been recruited. Participants were paired with individuals of the same gender forming two pairings: male/male (MM) and female/female (FF). Pairs did not know each other before the experiment.

Task and Procedures

The task used in the study is based on the social dilemma game used in Zheng's study [5]. Pairs of participants played a multi-trial variant of a day trader trust game. For this task, each participant was to imagine being a day-trader during a simulated multi-day (one month) investment period. In a single one hour session, the task simulated 30 "days" of investment with a large number of opportunities for partners to assess each other's behavior. After every five trials the participants were allowed to communicate via IM for up to 5 minutes, giving the participants the opportunity to communicate with the other trader in the game. During the discussion time, participants could talk about any aspect of the game with the other trader. In terms of the pre-task interaction, pairs in this condition had a 10-minute getting acquainted session, using IM to introduce themselves and get to know each other before they began the trust game.

Participants started with a background questionnaire, and then were paired to perform the day trader trust game either in the pre-task interaction condition or in the condition with no pre-task interaction. Afterwards, they completed the 8-item trust post-questionnaire adapted from McAllister [2]. The questionnaires used a 1-7 Likert-scale - 1 for strongly disagree and 7 for strongly agree.

Results

The primary results indicated that 1) gender as an important factor influences expectations and perceptions of communicators in CMC, in which female pairs had high levels of trust and more collaborative behaviors than male pairs in doing the task. 2) Females' collaborative conversational style focusing on harmonious relationships put them in a position to achieve trust in the communication. 3) Initial social chat prior to beginning work help remote team members build trust in the communication, particularly, initial social chat is more effective in female dominated work groups. 4) All pairs had better performance over time during the task. 5) Female pairs took less time in doing the task than male pairs. The results have implications for both the research and practice of building remote trust, as well as provide possible solutions to the design of interactive systems.

3 Significance to HCI

This study covers four areas of research, gender, CMC, trust and communication, and may impact society in several ways. First, this research theoretically offers good contributions to our understanding of the role of gender and initial social activities on trust, performance and behaviors as well as communication process. Although previous studies showed the importance of prior acquaintance and personal information on trust and cooperation, with detailed quantitative and qualitative analyses, our results provide additional insights on the impacts of both gender and initial social chat on trust and cooperation in more complex ways. It appears that initial social chats have larger benefit for females to build trust and behave highly collaboratively. When companies have to use low-end media, i.e., IM for remote communication, simple getting acquainted social chat before beginning work is expected to be a good solution for increasing trust perception and improving performance, especially for female dominated work groups. These companies include, but are not limited to, geographically dispersed virtual teams doing software development, business marketing, or customer service. Second, this research contributes to the literature of the effects of group gender composition on performance outcomes. To date, there is only a small body of literature that deals with this issue in HCI. Using multiple measures of cooperative behaviors, this research carefully investigates how different gender pairs behave in the trust game. The results bring practical implications to the work place. For instance, including females in the work group would possibly improve the work efficiency due to the fact that females tend to post relatively short messages, and are more likely to justify their assertions, apologize, and express support of others. Third, the discourse analysis of conversation style in IM provides further understanding of communication patterns in different gender groups, as well as provides more insights to understand that different levels of trust perceived by male and female gender pairs are partly due to the conversational styles they use. Today, a critical skill for managers is to become aware of the power of linguistic style and to make sure they understand the voices of all their employees. Indeed, as the work place becomes more diverse in term of gender culture and business becomes more global, team leaders will need to become even better at interacting with diverse team members and more flexible in adjusting their own styles to different group compositions.

1. Bos, N., Olson, J., Gergle, D., Olson, G., and Wight, Z. Effects of four computer-mediated communications channels on trust development. In Proc .of CHI2002, ACM Press (2002), 135-140
2. McAllister, D.J. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal* 38, 1 (1995), 24-59.
3. Spears, R., and Lea, M. Panacea or panopticon? The hidden power in computer-mediated communication. *Communication Research* (1994) 21, 427-459
4. Sun, X., Zhang, Q., Wiedenbeck, S., & Chintakovid, T. Gender Differences in Trust Perception when Using IM and Video. In Proc. of CHI2006, ACM Press, (2006)
5. Zheng, J., Veinott, E., Bos, N., Olson, J., and Olson, G. Trust without touch: jumpstarting long-distance trust with initial social activities. In Proc .of CHI2002, ACM Press (2002)

POSTER SUBMISSION 23:

Incentive Design for Home Computer Security

AUTHOR(S):

Wash, Rick

MacKie-Mason, Jeffrey King

ORGANIZATION(S):

University of Michigan, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Incentive Design for Home Computer Security

Poster Submission

People are the weakest link in security (Anderson, 1993). People write passwords on sticky notes on the screen. People don't patch their home systems and become botnet zombies. People choose whether to label a patch critical or just recommended. Our motivating insight is that these actions generally reflect *motivated behavior* in response to the configuration of incentives confronting individuals.

Since behavior is motivated by the goals and preferences of the individual, this behavior can be altered by designing appropriate incentives. By carefully structuring the benefits received from using a technology, we can induce users to make choices that most benefit the system. Along with some colleagues, we are developing a methodology for incentive-centered design of technology systems. We are working to provide guidelines and examples of how to carefully structure benefits to induce appropriate user choices.

We are applying these technology design ideas to a major open problem in computer security: botnets. Botnets are large collections of computers (called *zombies*) that are under the control of a single attacker. Botnets are behind a number of large security problems including spam email, distributed denial of service attacks, and multiple types of fraud and extortion (Ratliff, 2006). A significant part of the problem concerns security vulnerabilities inherent in the design of operating systems, network protocols and middleware. We do not address this well-studied issue. Instead, we focus on the problem that many zombies result from home computers that are poorly administered; that is, they are left more vulnerable than necessary given the current state of protective software. Home computer users frequently lack the skills necessary to properly secure their computer, and to properly clean the computer once it has been compromised. By providing appropriate incentives, it may be possible to induce these home users to make better choices in securing their computers.

An individual's use of software is largely driven by his or her perception of the direct benefits and costs of use (including the costs of learning the technology). The problems of non-use and mis-use are especially great for information security technologies for at least two reasons. First, many of the benefits of maintaining a secure system accrue not to the user, but to others. A home computer user rarely suffers from the insecurity he causes; it is the victims of the botnet that benefit from increased security. Ratliff (2006) describes how botnets can be used for extorting protection money from online businesses. Second, due to the nature of security systems, users are often not well-informed about the benefits to themselves. Most security systems are not directly productive; they exist to prevent productivity losses. As such, there is little feedback to users as to their own benefits (e.g. which losses were avoided) from their security choices. On the other hand, costs of recommended security behavior are usually more obvious, and thus receive more weight in user decisions. For example, CERT recommends¹ turning off Java and JavaScript, which

¹http://www.cert.org/tech_tips/home_networks.html

will cripple many popular websites such as Google GMail, MSN Games, and most so-called Web 2.0 services.

Integrating Ideas from Economics and Psychology

The botnet problem is not entirely unique, of course. It shares properties with a well studied problem in economics — the voluntary provision of public goods. A public good is any good that is both *non-excludable* and *non-rivalled*. A good is *non-excludable* if no one (in a well-defined group) can be prevented from benefiting from the good. A *non-rivalled* good can be used by many people at once, where one person using the good does not prevent others from also using the good. Public goods usually have a problem with free riding — people using the good without contributing to its creation. This problem occurs when the public good is provided through voluntary contributions. Individual people will rationally choose to contribute less than optimally. Andreoni (2006) and Chen (2003) provide good summaries of public goods research.

Thanks to botnets, home computer security is an instance of a public good. Increasing the security of a home computer increases the difficulty of forming a botnet, benefiting all of society. Conversely, a person can free ride by permitting his or her computer to remain insecure. Currently home computer security is voluntarily provided, and evidence indicates that home users are insufficiently securing their computers, leading to large numbers of easily-compromised zombies. This is a well-known observation, first observed by Camp and Wolfram (2000), and the literature is summarized by Anderson and Moore (2006) and Wash and MacKie-Mason (2007).

Psychology also has interesting theories to add to this problem. It identifies the problem as one of social loafing, which is the reduction in motivation when individuals work collectively. It is a problem of perception — individuals only social loaf when he or she perceives a group outcome rather than separate individual outcomes.

A meta-analysis (Karau and Williams, 1993) of the social loafing literature proposed a useful model that integrates the results of many experiments into a coherent theory. Basically, people are less motivated to work toward group goals when they cannot see as strong of a connection between their effort and the final outcome that they value. Therefore, social loafing can be reduced by increasing the perceived value and importance of the individual contributions. Ling et al. (2005) tested a number of useful design principles derived from this theory in the context of contributions to an online community.

Designing a Social Firewall System

We are using these ideas to design a new personal intrusion detection system, also known as a personal firewall. Personal firewalls monitor the local computer for events such as network accesses or applications being run. These systems then consult a policy to determine whether this event should be permitted to happen, or denied with an error. A good policy can limit the system to only be able to do what the user wants it to do, and nothing more. Most personal firewall systems have the ability to leave part (or most) of the policy incomplete, and prompt users interactively as needed. However, when prompted few users have the knowledge or experience necessary to make a 'good' choice. Our system tries to help users with this decision by sharing information from other users faced with a similar decision.

A simple mechanism for sharing information between users would simply share and aggregate policy decisions. Whenever an application attempts to access the Internet, the user will be presented with a dialog box asking them to “Permit” or “Deny” the access. This binary decision could then be sent to a central server to be aggregated with other people’s decisions. Future users, faced with the same decision, can see behavior of others (“80% of users have chosen ‘Permit’”) to help with their choice. This might work based on a ‘wisdom of crowds’ (Surowiecki, 2004) idea that while individual users are imperfect, the aggregate can make the correct choice. But is this what will happen?

We model the user’s decision in an economic model to attempt to predict user decisions. An appropriate model from economics is called ‘information cascades’ (Bikhchandani et al., 1998). We assume users prefer to make the ‘right’ choice between the two options, but there is uncertainty as to which choice is ‘right.’ The user also has some private information, a signal (like a gut feeling), that is accurate p percent of the time, where $p > 0.50$. This means that on average, the users’ gut feelings are correct, and if properly aggregated, the group could do better than if everyone chose individually. However, an important result was proved by Bikhchandani et al. (1992): there is a non-trivial probability that all users will choose incorrectly, despite receiving informative signals. This comes because users can only observe the results of other user’s discrete choices and cannot observe signals. If a relatively small number of users who make the first few choices have an erroneous signal, then all the subsequent users will rationally ignore their own signal and follow the previous users. Intuitively, if your gut says that a certain action should be ‘permitted’ but everyone before you has chosen ‘deny,’ then it is rational to question your gut and also choose ‘deny.’ However, by doing so you don’t provide any information about your ‘permit’ signal to future users. Future users encounter the same situation, observing everyone else choosing ‘deny,’ and will choose ‘deny’ for similar reasons. Anderson and Holt (1997) have validated this effect in the lab with human subjects.

We are exploring additional ideas for user contributions. It may be useful to have users report if they are compromised by a virus, worm, or hacker. Such occasional outcome information can help avoid the information cascade described in the previous paragraph. Additionally, it may be possible to collect comments from users describing why a given choice was made. These comments could be filtered, ordered, and voted upon much like reviews on amazon.com to provide useful information to other users. The difficulty in all these situations is motivating appropriate contributions by users. We hope to use design principles from the literature mentioned above from economics and psychology to motivate useful contributions.

References

- Lisa Anderson and Charles Holt. Information cascades in the laboratory. *American Economic Review*, 87(5):847–862, December 1997. URL [http://links.jstor.org/sici?sici=0002-8282\(199712\)87%3A5%3C847%3AICITL%3E2.0.CO%3B2-9](http://links.jstor.org/sici?sici=0002-8282(199712)87%3A5%3C847%3AICITL%3E2.0.CO%3B2-9).
- Ross Anderson. Why cryptosystems fail. In *CCS '93: Proceedings of the 1st ACM conference on Computer and communications security*, pages 215–227. ACM Press, 1993. ISBN 0897916298. doi: 10.1145/168588.168615. URL <http://portal.acm.org/citation.cfm?id=168615>.
- Ross Anderson and Tyler Moore. The economics of information security. *Science*, 314(5799):

- 610–613, October 2006. doi: 10.1126/science.1130992. URL <http://www.sciencemag.org/cgi/content/abstract/314/5799/610>.
- James Andreoni. Philanthropy. In S.-C. Kolm and J. Mercier Ythier, editors, *Handbook of Giving, Reciprocity and Altruism*, pages 1201–1269. North Holland, Amsterdam, 2006. URL <http://econ.ucsd.edu/~jandreoni/WorkingPapers/Philanthropy.pdf>.
- Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, and cultural change as information cascades. *Journal of Political Economy*, 100(5):992–1026, October 1992. URL <http://www.jstor.org/view/00223808/di980598/98p00557/0>.
- Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–170, Summer 1998. URL <http://www.jstor.org/view/08953309/di014715/01p0058j/0>.
- L. Jean Camp and Catherine Wolfram. Pricing security. In *Proceedings of the Information Survivability Workshop*, 2000. URL <http://www.springerlink.com/index/m44317165u727779.pdf>.
- Yan Chen. Incentive-compatible mechanisms for pure public goods: A survey of experimental research. In Charles Plott and Vernon Smith, editors, *The Handbook of Experimental Economics Results*, volume 1. Elsevier Science Publishing Company, April 2003. URL <http://www.si.umich.edu/~yanchen/papers/chenpub.doc>.
- Steven Karau and Kipling Williams. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65(4):681–706, 1993.
- K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, D. Cosley, D. Frankowski, L. Terveen, A. M. Rashid, P. Resnick, and R. Kraut. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4), 2005. URL <http://jcmc.indiana.edu/vol10/issue4/ling.html>.
- Evan Ratliff. The zombie hunters. *The New Yorker*, October 10 2006. URL http://www.newyorker.com/fact/content/articles/051010fa_fact.
- James Surowiecki. *Wisdom of Crowds*. Little Brown, 2004.
- Rick Wash and Jeffrey K. MacKie-Mason. Security when people matter: Structuring incentives for user behavior. In *International Conference on Electronic Commerce*, Minneapolis, MN, 2007. ACM. URL <http://www-personal.si.umich.edu/~rwash/pubs/icec702w-wash.pdf>.

POSTER SUBMISSION 24:

The Impacts of Gender and Initial Social Activities on Trust and Performance in CMC

AUTHOR(S):

Sun, Xiaoning

ORGANIZATION(S):

Drexel University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The Impacts of Gender and Initial Social Activities on Trust and Performance in CMC

1 Introduction and Research Questions

In recent years HCI researchers have made significant progress in understanding how different communication media influence people's trust perceptions and task performance in virtual environments [1, 5]. In this research little attention has been paid to the effects of gender differences in communication in virtual environments. However, there is good reason to investigate the role of gender, given research in the discipline of communication studies that has shown gender effects in face-to-face and virtual communications. In communication studies, gender analyses have focused on patterns of communication involving status, power, and influence [3]. By contrast, this research aims to bring the gender factor into computer-mediated communication (CMC) from an HCI perspective, responding to central concerns of efficiency, effectiveness, and user perceptions.

The primary objective of this research is to explore gender differences in synchronous computer-mediated communication with and without initial social activities. In particular, I ask whether initial social activities affect trust development and performance of male and female pairs in a social dilemma game.

To this end, the research questions for this study are:

- 1) How does gender influence people's trust development over multi-trials of a competitive task via the IM System?
- 2) How effectively does pre-task activity, i.e. social chat, help different gender pairings achieve higher levels of trust and better performance over multi-trials of a competitive task?
- 3) Is there a larger benefit of pre-task activity for males, who tend to be less trusting [4], than for females in doing a task which involves conflicts of interest? If the answer is YES, do males achieve the same level of trust as females?
- 4) Qualitatively, how do different gender pairings use language to communicate via the IM System and how does their language affect trust?

In order to get answers for these questions, I measure three dependent variables: participants' post-task trust, participants' performance and behaviors during the task and the pairs' communication processes.

2 Methodology and Primary Results

A 2x2 between subjects design was used: gender pairing (male/male vs. female/female), and pre-task interaction (interaction vs. no interaction). All participants used the AOL IM System to do the task. The pairs were strangers and the gender of the partners was revealed by the experimenter immediately before pairs performed their task.

Participants

A total of 124 participants have been recruited. Participants were paired with individuals of the same gender forming two pairings: male/male (MM) and female/female (FF). Pairs did not know each other before the experiment.

Task and Procedures

The task used in the study is based on the social dilemma game used in Zheng's study [5]. Pairs of participants played a multi-trial variant of a day trader trust game. For this task, each participant was to imagine being a day-trader during a simulated multi-day (one month) investment period. In a single one hour session, the task simulated 30 "days" of investment with a large number of opportunities for partners to assess each other's behavior. After every five trials the participants were allowed to communicate via IM for up to 5 minutes, giving the participants the opportunity to communicate with the other trader in the game. During the discussion time, participants could talk about any aspect of the game with the other trader. In terms of the pre-task interaction, pairs in this condition had a 10-minute getting acquainted session, using IM to introduce themselves and get to know each other before they began the trust game.

Participants started with a background questionnaire, and then were paired to perform the day trader trust game either in the pre-task interaction condition or in the condition with no pre-task interaction. Afterwards, they

completed the 8-item trust post-questionnaire adapted from McAllister [2]. The questionnaires used a 1-7 Likert-scale - 1 for strongly disagree and 7 for strongly agree.

Results

The primary results indicated that 1) gender as an important factor influences expectations and perceptions of communicators in CMC, in which female pairs had high levels of trust and more collaborative behaviors than male pairs in doing the task. 2) Females' collaborative conversational style focusing on harmonious relationships put them in a position to achieve trust in the communication. 3) Initial social chat prior to beginning work help remote team members build trust in the communication, particularly, initial social chat is more effective in female dominated work groups. 4) All pairs had better performance over time during the task. 5) Female pairs took less time in doing the task than male pairs. The results have implications for both the research and practice of building remote trust, as well as provide possible solutions to the design of interactive systems.

3 Significance to HCI

This study covers four areas of research, gender, CMC, trust and communication, and may impact society in several ways. First, this research theoretically offers good contributions to our understanding of the role of gender and initial social activities on trust, performance and behaviors as well as communication process. Although previous studies showed the importance of prior acquaintance and personal information on trust and cooperation, with detailed quantitative and qualitative analyses, our results provide additional insights on the impacts of both gender and initial social chat on trust and cooperation in more complex ways. It appears that initial social chats have larger benefit for females to build trust and behave highly collaboratively. When companies have to use low-end media, i.e., IM for remote communication, simple getting acquainted social chat before beginning work is expected to be a good solution for increasing trust perception and improving performance, especially for female dominated work groups. These companies include, but are not limited to, geographically dispersed virtual teams doing software development, business marketing, or customer service. Second, this research contributes to the literature of the effects of group gender composition on performance outcomes. To date, there is only a small body of literature that deals with this issue in HCI. Using multiple measures of cooperative behaviors, this research carefully investigates how different gender pairs behave in the trust game. The results bring practical implications to the work place. For instance, including females in the work group would possibly improve the work efficiency due to the fact that females tend to post relatively short messages, and are more likely to justify their assertions, apologize, and express support of others. Third, the discourse analysis of conversation style in IM provides further understanding of communication patterns in different gender groups, as well as provides more insights to understand that different levels of trust perceived by male and female gender pairs are partly due to the conversational styles they use. Today, a critical skill for managers is to become aware of the power of linguistic style and to make sure they understand the voices of all their employees. Indeed, as the work place becomes more diverse in term of gender culture and business becomes more global, team leaders will need to become even better at interacting with diverse team members and more flexible in adjusting their own styles to different group compositions.

1. Bos, N., Olson, J., Gergle, D., Olson, G., and Wight, Z. Effects of four computer-mediated communications channels on trust development. In Proc .of CHI2002, ACM Press (2002), 135-140
2. McAllister, D.J. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal* 38, 1 (1995), 24-59.
3. Spears, R., and Lea, M. Panacea or panopticon? The hidden power in computer-mediated communication. *Communication Research* (1994) 21, 427-459
4. Sun, X., Zhang, Q., Wiedenbeck, S., & Chintakovid, T. Gender Differences in Trust Perception when Using IM and Video. In Proc. of CHI2006, ACM Press, (2006)
5. Zheng, J., Veinott, E., Bos, N., Olson, J., and Olson, G. Trust without touch: jumpstarting long-distance trust with initial social activities. In Proc .of CHI2002, ACM Press (2002)



The Impacts of Gender and Initial Social Activities on Trust and Performance in CMC (#15)

Xiaoning Sun
Drexel University

Xiaoning.sun@school.drexel.edu

ABSTRACT

On carrying out complex, collaborative tasks in virtual environments, mediated by modern communication media. The primary objective of this research is to explore gender differences in synchronous computer-mediated communication (CMC) with and without initial social activities. In particular, it aims to investigate whether exposure to pre-task social activities before doing a task can help males, who tend to be less trusting as shown in previous research, overcome this barrier. Participants who did not previously know each other were recruited to form homogeneous pairs, male-male and female-female. Each pair carried out a competitive task via Instant Messenger either with or without pre-task social chat. The results from both quantitative and qualitative analyses suggest that gender as an important factor influences expectations and perceptions of communicators in CMC, in which female pairs had high levels of trust and more collaborative behaviors than male pairs in doing the task. In addition, females' collaborative conversational style focusing on harmonious relationships put them in a position to achieve trust in the communication. The results also suggest that initial social chat prior to beginning work help remote team members build trust in the communication, particularly, initial social chat is more effective in female dominated work groups. The results have implications for both the research and practice of building remote trust, as well as provide possible solutions to the design of interactive systems.

- ◆ Q1: How does gender information influence people's trust development and performance over multi-trials of a competitive task via the IM System?
- ◆ Q2: How effectively does pre-task activity, i.e. social chat, help male and female pairs achieve higher levels of trust and better performance over multi-trials of a competitive task?

METHODOLOGY

Participant & Task

- ◆ 20 male pairs (MM) and 33 female pairs (FF) undergraduate students. Participants in each pair didn't know each other beforehand.
- ◆ Gender of the partners was revealed by the experimenter immediately before pairs performed their task.
- ◆ A social dilemma game (Zheng, Winick, Bos, Olson, and Olson, 2002) was performed in the study.

Procedure and Measures

- ◆ Participants started with a background questionnaire, and then were paired to perform the day trader trust game either in the pre-task interaction condition or in the condition with no pre-task interaction. Afterwards, they completed the 8-item trust post-questionnaire adapted from (McAlister, 1995). The questionnaires used a 1-7 Likert scale - 1 for strongly disagree and 7 for strongly agree.
- ◆ Four measures: trust perception from questionnaire responses, performance time, cooperative behaviors during the task and qualitative analysis on communication process.

DISCUSSION AND CONCLUSION

- ◆ did not
- ◆ Female pairs had higher number in which they invested fully ...
- ◆ Females tended to post short time
- ◆ Discussion on Q2
- ◆ The effect of prior acquaintance cooperation (Rocco, 1996; Zheng
- ◆ People who meet before working relationship
- ◆ The effects of repeated interaction
- ◆ Importance of the direct interaction
- ◆ Discussion on Q3
- ◆ NO
- ◆ larger benefit of pre-task
- ◆ Other possible ways to improve trust: low levels of trust
- ◆ longer time of pre-task interaction?
- ◆ other type of media, e.g., video
- ◆ Discussion on Q4
- ◆ Female pairs used more statements
- ◆ cues of group payoff or their
- ◆ Male pairs used more direct action, such as strategy and
- ◆ Female pairs used more emotions
- ◆ (to be confirmed) affects the communication

Fig. 1 Participant perception

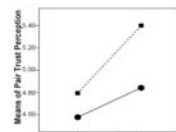
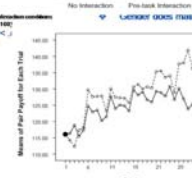
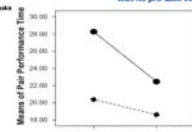
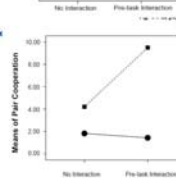


Fig. 2 Pair cooperation



POSTER SUBMISSION 25:

Federal Funding for Information Research

AUTHOR(S):

Hahn, Trudi Bellardo

ORGANIZATION(S):

University of Maryland, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Based on a literature search, interviews with program officers at federal funding agencies, and other documentation, this poster addresses the patterns of funding for information science research, focusing specifically on faculty in iSchools. The study examines several major funding agencies and their missions and funding priorities, and concludes with recommendations for increasing the chances of success in winning federal grants.

Questions addressed in this study:

Which agencies have contributed the most support?

What initiatives have been supported, how has the emphasis on topics and issues changed, where are there gaps, and which initiatives are currently at the top of federal agencies' research agendas? How has the amount of funding fluctuated over the years, what factors have influenced the availability of federal dollars for LIS research, and what is the prognosis for the future?

How are research agendas currently being set? Who is determining priorities?

These and other questions are addressed with charts, tables, and other graphical displays showing trends and current priorities, which will invite questions and informal discussion.

POSTER SUBMISSION 26:

The Collaborative Information Behavior of Middle School Students in Online Learning Environments: An Exploratory Study

AUTHOR(S):

Zhou, Nan

Agosto, Denise

ORGANIZATION(S):

Drexel University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The Collaborative Information Behavior of Middle School Students in Online Learning Environments: An Exploratory Study

Introduction

The emergence of social computing systems such as blogs and online communities is shaping people's way of working, learning, and interacting with others. Teenagers and preteenagers are leading the transition "to a fully wired and mobile nation" (Lenhart, Madden, & Hitlin, 2005). A growing body of research indicates that today's young people, often referred to as the Millennial Generation or as "digital natives" (Prensky, 2001), are heavy users of social computing media and that much of their information seeking and socialization takes place via Instant Messaging (IM), text messaging, blogs, wikis, email, and the World Wide Web (DeBell & Chapman, 2006; Harris, 2005; Lenhart, Madden, & Hitlin, 2005; Rideout, Roberts, & Foehr, 2005). Although the study of human information behavior (IB) is a rapidly growing subset of library and information science research, most IB research has focused on studying individual behaviors. We need to understand better how people "need, seek, give, and use information" (Fisher et al, 2006) *in groups* and in the context of online communities that use social computing technologies as well.

At the Virtual Math Teams (VMT) project (<http://mathforum.org/vmt/>), we are studying how middle school students work on math problems together and how they create and maintain online communities. As part of the VMT's larger research agenda, this study explores how students collaborate on math problem solving in online learning environments and looks into how they identify and address their information needs during this process.

Research Method

The VMT project has held over 80 sessions with groups of U.S. middle school students in naturalistic settings where they log into the online environments from home or school. Interactions are mediated entirely through the virtual environment using VMT Chat. Developed specifically for the project, VMT Chat includes a chat program and a shared whiteboard with drawing tools. Typical VMT sessions are about one hour long and are recorded for later participant access as well as for analysis purposes.

Taking the group as the unit of analysis, we have adopted an ethnomethodologically-informed approach (Garfinkel, 1967) based on aspects of conversation analysis (Sacks, 1992) to analyze participant interactions. We have held weekly data sessions at VMT during which researchers examine data excerpts of interest and discuss observations and findings. Such micro-level analysis usually focuses on episodes of activities that are of research interest and tries to identify and describe the observable methods participants use to make sense of their interactions from an interactional perspective.

Findings

Constructing shared meaning

As is typical of the Millennial Generation, most of our participants are Internet savvy and are skillful users of online chat programs. They bring the behaviors they are familiar with into the online environment and use various methods either that they already knew of or that they construct using available resources to make sense of the new environment and the interactions taking place within it. At the same time, they may not be experienced with collaborating online, especially in doing mathematics, since math is usually considered a solitary endeavor in classroom settings. The chat environment is a complex social world where participants come with certain goals and expectations. Together as a group, they coordinate with each other and construct their shared experience of collaborating online. Among all sorts of activities participants

tend to engage in, one noticeably important behavior during the collaboration process involves working together to achieve shared meaning. The work participants do to construct shared meaning is not merely simple information exchange but rather co-construction of meaning, which emerges as an achievement of the interactions.

Information needs as an interactional achievement

Provided the *math problem* they are to work out together, students first need to construct *their own problem*. That is, they need to identify what is known and what still needs to be figured out. This process is negotiated and the problem is constituted interactionally. Participants ask various questions to clarify, to request explanation or elaboration from their teammates, to seek information, to engage others to participate, and so on. Information needs for math problem solving emerge in and through the interactions as an interactional achievement. There are observable members' methods (Garfinkel, 1967) and tactics participants use for doing such collaborative formulation of information needs. Under such situations, information needs often emerge 1) when there is imbalance of what is known among the group members and 2) when the group identifies a shared knowledge deficiency. Once an information need is identified, participants use different methods to satisfy it, for example, by offering what they know, seeking information from the group, or using external resources.

Group as information source

The data from the VMT sessions also show that participants take the group as one primary resource for information. There are different ways questions are posed within the group. Some are designed and to be read as calling for reminder of math information when one claims that he "forgot." Some call for factual information, whereas some ask for explanation of work that has been done by other members. Often, participants provide information in order to seek information. They also may provide possible answer to an information question. Information is also provided by participants in different ways when responding to questions posed within the group. In some cases, one participant provides information intended to directly answer a question from another group member, while in other cases a reference is provided that points to some online resource that presumably has information to answer the question.

Implications

This study is a first step in understanding collaboration and learning in online chat environments. An understanding of student collaboration in social computing environments such as the VMT will go a long way toward building a better understanding of the information behaviors of the Millennial Generation, a generation that seems to place strong emphasis on group interaction and learning (Oblinger, 2003). Such an understanding could also lead to recommendations for designing online learning environments and digital libraries (Zhou & Stahl, 2007) that can better support students' group learning behaviors and preferences.

References

- DeBell, M. & Chapman, C. (2006). *Computer and internet use by students in 2003: Statistical analysis report*. Washington, D.C.: U.S. Department of Education.
- Fisher, K. E., Erdelez, S., & McKechnie, E. F. (Eds.). *Theories of information behavior*. Medford, NJ: Information Today.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Harris, F. J. (2005). *I found it on the Internet: Coming of age online*. Chicago: American Library Association.

- Lenhart, A., Madden, M., & Hitlin, Paul. (2005) *Teens and technology: Youth are leading the transition to a fully wired and mobile nation*. The Pew Charitable Trusts. Retrieved December 28, 2005 from http://www.pewinternet.org/PPF/r/162/report_display.asp.
- Oblinger, D. G. (2003). Boomers, Gen-Xers, and Millennials: Understanding the new students." *Educause*, 38(4), 36-47.
- Rideout, V., Roberts, D. F., & Foehr, U. G. (2005). *Generation M: Media in the lives of 8-18 year-olds*. New York: Kaiser Family Foundation.
- Sacks, H. (1992). *Lectures on conversation*. Oxford, UK: Blackwell.
- Zhou, N., & Stahl, G. (2007). Information behavior of small groups: Implications for design of digital libraries. Paper presented at *the Joint Conference of Digital Libraries (JCDL '07)*, Vancouver, Canada.

POSTER SUBMISSION 27:

Watching Organizational Opinion via Social Tagging

AUTHOR(S):
Russell, Terrell

ORGANIZATION(S):

University of North Carolina-Chapel Hill, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Our collective understanding of the world around us is largely constructed upon what others think. We trust our neighbors more than strangers. We know things because we have either seen them ourselves or because we have it "on good information".

On an individual level, the process of keeping track of what your neighbors' opinions are has largely been a cognitive exercise. Few of us took the time to document the opinions of our peers and acquaintances - much less analyze that documentation for patterns or correlations. With the advent of social tagging, this collective opinion can be inexpensively visualized and reflected upon, counted and analyzed.

Tagging data can be viewed as a marker for the pulse of an organization. With the proper visualizations and training, social tagging could become the most powerful information organization and management tool available.

The tagging triumvirate consists of the person doing the tagging, the term being applied, and the object being tagged. What is usually forgotten is the fact that the tagging event happens at a certain time. Tracking how the tagging practices of an organization or group change over time can deliver valuable insight into what the organization is doing and thinking.

If the object being tagged is explicitly restricted to someone's performance or expertise, a very interesting dataset could be collected quickly. What a group thinks someone is good at, while subjective, could be counted.

Perhaps more interesting is the potential for seeing the time-lapsed tagging behavior around a collection of people's expertise. Entire groups of talent could be analyzed as a whole with careful consideration given to the points in time where learning and understanding were apparently happening.

Building on earlier work from Cloudalicio.us, this visualization technique could afford an organization new insight into how it understands its own membership.

POSTER SUBMISSION 28:

Stability and change in self-organizing technology-supported groups: Evidence from genres of communication in free and open source software development

AUTHOR(S):

Howison, James

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Stability and change in self-organizing technology-supported groups
Evidence from genres of communication in free and open source
software development

(Author suppressed for review)

29 October 2007

Introduction

New organizations supported by information and communications technologies, such as open source software development and Wikipedia, have recently emerged with great and somewhat surprising success. These organizations are emergent, self-organizing and appear able to change and adapt as success breeds new challenges, all without undermining the creative—even emancipatory—collaborative experience.

This abstract proposes a poster presentation for the iConference 2008, drawing on dissertation work in progress with an intended completion date of April 2008. Therefore the poster would present the conceptual framework, method and preliminary results at a time when the author can still incorporate feedback from the iSchool community.

The dissertation work described in this proposal will further our understanding of the emergence, stability and change in the practices of self-organization, through a systematic historical study of two comparable Free and Open Source Software development projects. Specifically it examines emergence, stabilization and change in the genres and genre-systems the projects use to communicate.

The research questions

1. What are the organizational practices of these FLOSS projects?
2. In what way are these practices linked to effectiveness over time?
3. How and why do these practices emerge, stabilize and change?

The importance of the study

This study is important and timely in three ways. Firstly self-organizing teams, supported by information and communications technologies, are rising in importance throughout organizations as their number and areas of involvement rise. Secondly, the study of organizational change (and stability) is independently important as organizations attempt to improve in rapidly changing contexts. Together the study of organizational change in self-organizing teams provides the right context to examine a key theoretical question in the study of organizations, whether (or under what circumstances) organizations are essentially stable, with short periods of change, or essentially always in flux, with periods of seeming stability disguising continuous, incremental change (Orlikowski, 1996; Pettigrew et al., 2001). Finally Free (Libre) and Open Source software (FLOSS) development is not only important in itself but is also a canonical and useful domain to study the issues outlined above.

Conceptual Framework and Literature Review

In order to empirically answer these research questions a solid, operationizable conceptual framework, or theoretical base, is needed. The conceptual framework must be able to a) conceptualize organizational practices so that stability and change can be observed, b) understand effectiveness in the FLOSS context. and c) explain both 'how' and 'why' stability or change occurs.

The genre approach to organizational practices

Yates and Orlikowski define genres of organizational communication as “socially recognized types of communicative actions—such as memos, meetings, expense forms, training seminars—that are habitually enacted by members of a community to realize particular social purposes” (Orlikowski and Yates, 1994, p 54). They are a type of institution, which Barley and Tolbert (1997) identify as “historical accretions of part practices and understandings that set conditions on action” (p. 99). Genres provide an operationalization of organizational routines (Feldman and Pentland, 2003) appropriate for the distributed team context.

Every domain of activity has characteristic patterns of communication. In the academic world the submission of a research paper follows a characteristic sequence of documents from many participants, from blinded submissions, reviews, editor’s letters, response to reviewers and acceptance or rejection letter. The application of the genre approach to the FLOSS context proposes that the communications that FLOSS members craft also follow particular patterns seen as legitimate by teams.

Orlikowski and Yates (1994, p. 54) argue that genre, and genre systems, can be discovered by observing “why, what, who/in, how, when and where” of communications and sequences; the features of genre are purpose, content, participants, form, rhythm and timing, and venue. Together “the set of genres that are routinely enacted by members of the community” form a “genre repertoire.” Stability and change is possible to identify because, with historical detail extracted from archives as described below, a catalogue of genres in use at different times will reveal change, both in the relative use of different genres but also within genres (and genre systems), as features are increasingly or decreasingly present.

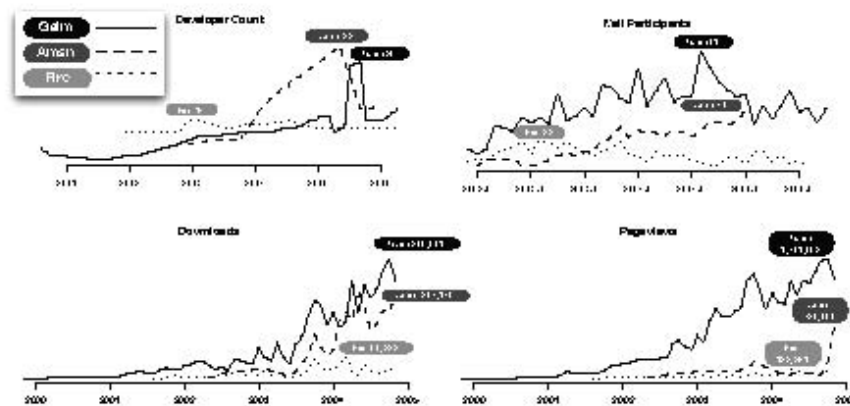
Emergence, stability and change

The genre approach provides an appropriate conceptualization of practices and an explanation of 'how' they both achieve some stability (through structural observation and repetition) and change (through importation and improvisation). However the genre approach has to date only discussed 'immediate' causes, or perhaps opportunities, for genre change, drawing on Gersick and Hackman (1990). However, particularly in a self-organizing group, deeper explanations of the underlying drivers of these opportunities are desirable, allowing the research to explain, for example, why novelty is encountered or milestones chosen or the structure of the group altered.

While there is not sufficient space here to describe it the literature touching on change in the FLOSS context does suggest a generative model of change (Senyard and Michlmayr, 2004; Oh and Jeon, 2004; Koch and Schneider, 2002; Robles et al., 2005; O'Mahony, 2003), which draws on the useful taxonomy of organizational change "generative mechanisms" provided by Poole and Van de Ven (2004).

These studies suggest a compound generative mechanism for change which turns on a dialectical tension between the desire to attract participants and the need to maintain cohesion in the team. The lifecycle effects of code growth and audience diversification eventually tip the balance so that the key organizational challenge shifts from inclusion to exclusion. Such changes ought to be visible through the method below as certain genres 'wax and wane'. Failure to adequately adjust might be the cause of reduced effectiveness. It should be noted, however, that this research is not designed to test this specific model directly. Different models of generative change might emerge from the method described below, but they would be similarly conceptualized and supported with evidence of trends in practices and effectiveness.

Figure 1: *Gain and Fire differ in their ability to sustain early success*



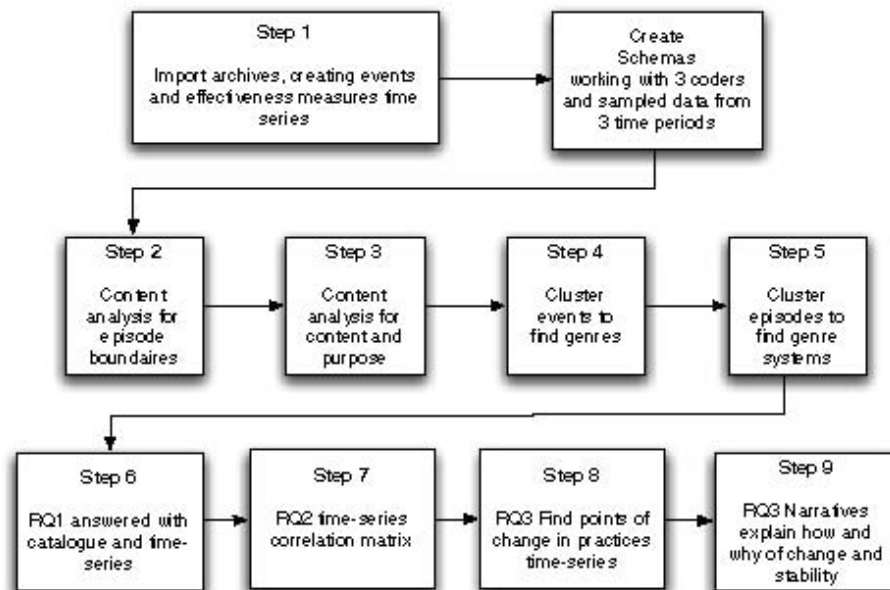
Conceptualizing effectiveness

The third research question links practices and effectiveness over time. The FLOSS context provides a challenge for creating appropriate measures of effectiveness. The multiple measures chosen for this study reflect the four categories suggested by Crowston et al. (2003, 2006, 2006): a) system creation and maintenance (such as code growth and developer attraction), b) system quality (such as complexity metrics), c) system use (such as downloads and pageviews) and d) system consequences (such as competitor reactions). Figure 1 shows sample time-series for comparable projects.

Proposed Method

It is proposed to pursue the research questions above through a comparative case study of two comparable FLOSS instant messaging projects, Gain and Fire, which differ in their ability to sustain their early success; Gain has gone from strength to strength, while Fire eventually stalled. An outline of the overall method is shown in Figure 2.

Figure 2: The proposed steps of this research



Two content analytic schemes are being developed. The first allows the recognition of *episodes* in the data: sets of communications linked by topic and shared attention. The second scheme describes the *purpose* of the emails, based on speech acts theory (Goldstein and Sabin, 2006; Searle, 1969), as well as the *content* (a ground up literal description, such as “description of application behavior”). These codes are combined with automatically genre attributes (“where, how, when and who”) to describe each communication on the six dimensions of genre. Clustering techniques will then be used to identify repeated patterns and these clusters will be interpreted as genres. Genre systems will be identified through clustering based on sequence analysis of genres in episodes (Bakeman and Gottman, 1997; Abbott, 1995, Abbott, 1995).

The question of practices in use, RQ1, will be answered with a typology of genres and

genre-systems that make up the project's "genre repertoire", together with a deeply linked catalogue of examples. This catalogue can be presented over time using both absolute and relative frequency to see how genre and genre-system use has changed.

RQ2 will be answered by regressing these time-series against the time-series of effectiveness measures. The effectiveness time-series may also indicate 'breaks' which show rapidly rising or falling effectiveness and such 'breaks' can be compared with changes in genre, or genre features, around those periods. For example it would be possible to test whether an observed lengthening of particular genres (perhaps discussions about including new features, or interminable procedural discussion) was associated with effectiveness trends such as declining developer participation.

The question of emergence, stability and change, RQ3, will be answered in two ways, one quantitative and one qualitative. Quantitative indicators of stability and change will be discovered by analyzing the genre and genre-system time-series to discover periods with statistically significant 'breaks', using interrupted time-series experiment techniques (Hamilton, 1994). The question of continuous vs punctuated change ought to be visible in the smoothness of the time-series, especially those graphing features of particular genres.

Such 'breaks' will indicate periods for further qualitative investigation producing micro-narratives of the emergence and change of particular genres, and macro-narratives of change in the projects, drawing on evidence from the relative frequency of particular genres and genre-systems. For example intentional change may be explicitly discussed and planned just prior to a 'break.' Or the arrival of a new genre may be precipitated by the arrival of a new participant particularly associated with it, indicating importation. The emergence of quality assurance messages may be accompanied by the emergence of an audience for stability.

Conclusion

The research will further our understanding of change in self-organizing groups as well as the 'via document' collaboration that increasingly constitutes work today.

Bibliography

- Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology*, 21:93–113.
- Bakeman, R. and Gottman, J. M. (1997). *Observing interaction : an introduction to sequential analysis*. Cambridge University Press, New York, 2nd ed edition.
- Barley, S. R. and Tolbert, P. S. (1997). Institutionalization and structuration: Studying the links between action and institution. *Organization Studies*, 18(1):93–117.
- Crowston, K., Annabi, H., and Howison, J. (2003). Defining open source software project success. In *Proc. of International Conference on Information Systems (ICIS 2003)*.
- Crowston, K., Howison, J., and Annabi, H. (2006). Information systems success in free and open source software development: Theory and measures. *Software Process: Improvement and Practice*, 11(2):123–148.
- Feldman, M. S. and Pentland, B. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48(1):94–118.
- Gersick, C. J. G. and Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes*, 47:65–97.
- Goldstein, J. and Sabin, R. E. (2006). Using speech acts to categorize email and identify email genres. In *HICSS '06: Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, page 50.2, Washington, DC, USA. IEEE Computer Society.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press, Princeton, N.J.
- Koch, S. and Schneider, G. (2002). Effort, cooperation and coordination in an open source software project: Gnome. *Information Systems Journal*, 12(1):27–42.
- Oh, W. and Jeon, S. (2004). Membership dynamics and network stability in the open-source community: the Ising perspective. In *Proceedings of International Conference on Information Systems (ICIS) 2004*.
- O'Mahony, S. (2003). Guarding the commons: how community managed software projects protect their work. *Research Policy*, 32(7):1179–1198.
- Orlikowski, W. J. (1996). Improvising organizational transformation over time: A situated change perspective. *Information Systems Research*, 7(1):63–92.
- Orlikowski, W. J. and Yates, J. (1994). Genre repertoire—the structuring of communicative practices in organizations. *Administrative Science Quarterly*, 39(4):541–574.
- Pettigrew, A. M., Woodman, R. W., and Cameron, K. S. (2001). Studying organizational change and development: Challenges for future research. *Academy of Management Journal*, 44(4):697–713.

- Poole, M. S. and Van de Ven, A. H. (2004). Theories of organizational change and innovation processes. In Poole, M. S. and Van de Ven, A. H., editors, *Handbook of Organizational Change and Innovation*, pages 374–397. Oxford University Press, New York, NY.
- Robles, G., Amor, J. J., González-Barahona, J. M., and Herraiz, I. (2005). Evolution and growth in large libre software projects. In *The 8th International Workshop on Principles of Software Evolution*, Lisbon, Portugal.
- Searle, J. (1969). *Speech Acts*. Cambridge University Press, Cambridge, UK.
- Senyard, A. and Michlmayr, M. (2004). How to have a successful free software project. In *Proceedings of the 11th Asia-Pacific Software Engineering Conference*, pages 84–91, Busan, Korea. IEEE Computer Society.

POSTER SUBMISSION 29:

“Community Innovativeness”: A New Perspective on Knowledge Creation

AUTHOR(S):

Addom, Benjamin

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Community Informatics (CI) is an emerging field of study that concerns itself with the design and use of Information and Communications Technologies (ICTs) to enable and empower local communities (Gurstein, 2000). But while the intention of designing appropriate technologies to meet the needs of these communities may sound very ethical, the approach oftentimes lacks rigor. One of the key areas lacking attention is the relationship between community structure and knowledge generation. This lack of adequate work has led to the assumption that communities are social systems that need to adopt outside innovations developed by people in white coats. The problem is exacerbated with the advent of the new ICTs that are seen as panacea to community development issues. This poster presents a research proposal that values a deeper awareness of community structure as key resource for design and use of ICTs for community development. The study is interested in identifying the underlying structural properties and processes that motivate communities to take initiatives in generation new knowledge. The main question, however is, how best can researchers, practitioners and policy makers (interested in the application of ICTs for community development) approach this process?

While the concept of “community” is still a site of contestation in general and specifically in CI, this study takes a geographic or spatial approach to community. The study also approaches innovation from user perspective (von Hippel, 2001) and identifies knowledge generation as key in community innovativeness. In other words, community innovation is a social process that leads to knowledge creation. It is the interplay between tacit and explicit knowledge. The study also recognizes the transformative power of the Internet in disseminating global codified knowledge to the developing and transitional countries (Stiglitz, 1999), but skeptical about its ability to unearth the “tacit” or “sticky” (Szulanski, 2003; Sevigny and Prevost, 2006) component of knowledge in these communities. Consequently, for the transformative power of the Internet to be evidenced in developing economies, there is the need to explore ways by which this sticky knowledge can be tapped and combine with the “transferred” codified knowledge to the benefit of its people.

This study argues that the structure of a given community is a predictor of the innovativeness of that community. Community Innovativeness is the dynamic capability of a relatively stable geographic system consisting of networks of social organization interacting significantly with each other in an institutional setting to acquire, assimilate, transform and exploit knowledge to meet the needs of the system. The model is driven by the theory of absorptive capacity (Cohen & Levinthal, 1990; Zahra & George, 2000); community development theory (Warren, 1978); community integration (Coleman, 1957; Aiken & Alford, 1970); theories of communication network (Monge & Contractor, 2003); and the dynamic theory of knowledge creation (Nonaka & Takeuchi, 1995; Szulanski, 2003).

"See attached file for reference"

POSTER SUBMISSION 30:

Beyond the Digital Divide into Computer-Mediated Communications: A Content Analysis of the Role of Community Weblogs in Building Oldenburg's Virtual Third Places in Black America

AUTHOR(S):

Igwe, Frank

ORGANIZATION(S):

Pennsylvania State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Beyond the Digital Divide into Computer-Mediated Communications: A Content Analysis of the Role of Community Weblogs in building Oldenburg's Virtual “Third Places” in Black America

Introduction

Although times change, there are certain human elements that survive through the ages. These elements include the need for expression, companionship, involvement, connection and information. The avenues by which humans engage in these social practices have evolved, and with the dawn of the Information Age we are seeing the emergence of new forms of computer mediated communication (CMC), with weblogs (or blogs) being a manifestation of this transformation. Some argue that within the on going information age in American society, no single individual is disconnected or separated from others due to the existence and prevalence of Information Communicative Technologies (ICT). These technologies have proved to be indispensable and are critical to communication between individuals, cultures and communities within the larger American society (Milne 2004).

This paper deals with these Information and Communicative Technologies, more specifically how weblogs (or blogs) are used by African Americans on the positive side of the digital divide to participate in virtual communities to rebuild communicative aspects of community that have been lost in “real-world” communities. In addition, the paper will determine whether these virtual communities fit the description of “third places” as portrayed by Oldenburg (Oldenburg 1999). These “third places” arise out of a need for individuals to find a dependable, neutral place of refuge to gather and interact, away from first places (home) and second places (work), often conferring or dealing with issues that may considered too taboo for public discussion by the community at large.

With this in mind the researcher identified an issue within the African American community that was of consequence, and yet was not being addressed due to individual or social pressures. The problem that presented itself was the lack of discussion and social support pertaining to the Human Immunodeficiency Virus (HIV), and Acquired Immune Deficiency Syndrome (AIDS).

African American HIV/AIDS Statistics

HIV/AIDS statistics paint a particularly disturbing picture for African American females, due to the fact that they account for a disproportionate number of infections relative to other social groups (Phillips 2005), and 75% of new HIV/AIDS cases within the larger African American population. The Center for Disease Control (CDC), states that HIV/AIDS is among the top 4 causes of death for African American women aged 25–54 years, and the number 1 cause of death for African American women aged 25–34 years (CDC 2006). In 2001, HIV/AIDS was among the top three causes of death for African American men 25–54 years of age, and of persons diagnosed with AIDS since 1995, a smaller percentage of African Americans (60%) were alive after nine years compared with whites (70%) (due in part to late diagnosis) (Health 2007). Despite these figures, there is still a deafening silence associated with the discussion of the disease, because contraction of the HIV virus is seen as a consequence of behaviors that are stigmatized within the largely religious and conservative African American community (i.e., promiscuity, homosexuality, or drug use), framed within the context of sin and immorality (Baker 1999).

Communities and Expression

An interesting element of any functional community is that it is self-sustaining. In order to be self-sustaining a community has to possess the ability to address issues that affect members' wellbeing, in either a direct or indirect fashion, to ensure that what members are getting out of the association exceeds the cost. Every healthy community discusses issues that threaten its survival. However, the number of African Americans infected and dying from HIV/AIDS is staggering, and the silence associated with the epidemic is akin to having “an elephant in the room” that nobody wants to talk about. This conflict, and the fear of violating group discussion norms, has created a prevalent silence on the subject, and degraded aspects of community, namely: emotional safety, sense of belonging, and positive reinforcement found in dialogue.

It is believed that in an effort to “heal” itself of this silence, and restore aforementioned communicative elements of community that have been diminished, African Americans have resorted to finding other outlets to discuss the epidemic. One outlet is believed to be found in online third places. Blogs were chosen because they represent a single place, outside of large social gatherings, where people can engage in real time conversations on a grand scale, and unlike their physical counterparts, users are empowered by the relative cloak of anonymity afforded by the internet. Never before has a medium such as ICT existed that can connect and enable conversations from members representing all classes of the community, with potentially everyone able to contribute to the discussion

and be heard. By technology being an enabler for rebuilding aspects of community, it adds impetus to the drive towards eliminating the “digital-divide” through tangible benefits, such as improved health outcomes through preventative, rather than reactive, practices.

Oldenburg’s “Third Places”

The silence associated with HIV/AIDS by traditional African American institutions created a need to form social bonds in other arenas. This need provided fertile conditions for the rise of an alternate place to gather and discuss transcendent issues; such places are typically born out of a need, and can be labeled as “third places”. Every stable community is comprised of a first place (home), a second place (work), and a third place (informal gathering location). If one of these components is missing, it affects the stability of the community (Baker-Eveleth, Eveleth et al. 2005). The third place (bars, cafes, barbershops, etc.) provides a context for sociability, spontaneity, community building and emotional expressiveness” (Oldenburg and Brissett 1982).

Ray Oldenburg states that third places are sought out by individuals because the function it serves is missing from their lives; more specifically, a lack of community poisons an individual’s sense of well being (Baker-Eveleth, Eveleth et al. 2005), and people will seek an exogenous source to fill that void if needed. Within the framework of computer mediated communication there has been a realization that cyberspace (such as blogs, chatrooms, etc.) resemble types of social settings described by Oldenburg, providing an informal place where individuals gather to rebuild communicative aspects of community that may be lost (Soukup 2006).

The Nature of Third Places

The most important characteristic of third places is that they unite a neighborhood, and act as a “mixer”. They bring together a wide range of individuals from various backgrounds, and within this mix individuals find others who share their same interest, or find others who do not share their interests but are still considered interesting, and viable candidates for future associations later on. Third places are places where members discuss politics, philosophy, geography, urban development, psychology, history and many other topics, and these dialogs offer an opportunity to flesh questions out, sound out protests, and to form opinions (Baker-Eveleth, Eveleth et al. 2005).

Third places are a political force to be reckoned with, and in times of crisis or of social upheavals, the seeds are often planted and watered in third place locales. Examples include the anti-labor segregation laws in the 1960’s, which were the result of assembly of blacks in churches all over the south. Other examples include labor solidarity which emerged in many communities after workers met in local cafés, discussed their common problems, and realized their collective strength, which enabled them to effectively plan strikes and other strategies (Oldenburg 1999). Oldenburg describes the essential characteristics of third places as (Oldenburg 1999; Soukup 2006)²:

1. Being on neutral ground
2. Being levelers
3. Conversation being the main activity, with the mood being playful
4. Accessible
5. Are a home away from home, and have “regulars”

African American Virtual Third Places

Research has found that there is indeed a vibrant community of Black bloggers (Poole 2005) that is known as the “Blackosphere”, as described by Francis Hollander (2007):

These blogs are by and principally for Black people, focusing not only upon Black people but upon people and issues deemed relevant to the Black people who write these blogs and post comments. At Black blogs, we comment on the issues of the day raised in white newspapers and blogs, but we also highlight issues that whites mostly ignore.

² See Oldenburg (1999) for a comprehensive review of Third Places

The research in this paper focuses on one of these sites found in the blackosphere, and addresses a basic question:

How, and if so why, does the Weblog under study exhibit characteristics of “third places” as conceptualized by Oldenburg (1999)?

Preliminary empirical findings from research of this “Blackosphere” reveals that there are indeed a significant number of individuals within the African American community that utilize blog sites to discuss the crisis of HIV/AIDS within the African American community. These blogs are a place where rich conversations take place, which helps increase community cohesiveness and emotional wellbeing as participants realize that they were not alone in dealing with the ramifications of the disease running rampant in the community (Igwe 2007 Forthcoming). The research lends credence to the belief that the importance of studying technology lies in the uniqueness of the social interactions the internet inspires, and not strictly for its technological attributes.

References

Baker, S. (1999). "HIV/AIDS, Nurses, and the Black Church: A Case Study." *Journal of the Association of Nurses in AIDS care* 10(5): 71-79.

Baker-Eveleth, L., D. M. Eveleth, et al. (2005). "An Emergining On-Line "Third Place" For Information Systems (IS) Students: Some Preliminary Observations." *Journal of Information Systems Education* 16(4).

CDC, D. o. H. a. H. S. C. f. D. C. a. P. (2006). Fact Sheet.

Health, N. I. o. (2007). Health Disparities in HIV/AIDS: Focus on African Americans (R03).

Milne, J. M. (2004). Weblogs and the Technology Lifecycle: Context, Geek-Chic and Personal Community. Department of Anthropology, University of South Florida.

Oldenburg, R. (1999). *The Great Good Place: Cafes, Coffee Shops, Bookstores, Bars, Hair Salons and Other Hangouts at the Heart of a Community*. New York, Marlowe & Company.

Oldenburg, R. and D. Brissett (1982). "The Third Place." *Qualitative Sociology* 5(4).

Phillips, L. (2005). "Deconstructing "Down Low" Discourse: The Politics of Sexuality, Gender, Race, AIDS, and Anxiety." *Journal of African American Studies* 9(2): 3-15.

Poole, A. (2005). *Black Bloggers and the Blogosphere*. The Second International Conference on Technology, Knowledge and Society, Hyderabad, India.

Soukup, C. (2006). "Computer-mediated communication as a virtual third place: building Oldenburg's great good places on the world wide web." *New Media & Society* 8(3).

Roundtable Discussions

(14 Accepted Submissions)

ROUNDTABLE SUBMISSION 1:

Beyond the qualitative/quantitative split: alternate forms of research in the information space

AUTHOR(S):

Feinberg, Melanie

Nathan, Lisa

Turner, Deborah

Unsworth, Kris

ORGANIZATION(S):

University of Washington, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The iSchool research space has welcomed mutlidisciplinary. Yet, research methods have most typically meant the extension of “information science” to include social science, and the use of qualitative as well as quantitative research methods. This session will permit attendees to consider implications mutlidisciplinary has on research methods. What opportunities may be afforded by more fully integrating a diverse range of humanities, design, and the arts methods that at times eschew goals of generalizability and predictability and place greater emphasis on the creativity of the researcher? Brief descriptions of four research projects incorporating less traditional methods will serve as case studies to ground and propel this discussion.

ROUNDTABLE SUBMISSION 2:

I Schools in the Greater Community

AUTHOR(S):

Kansa, Eric Christopher

Wilde, Erik

Yee, Raymond

ORGANIZATION(S):

UC Berkeley, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Information Schools have much to offer the greater university community. The organization, management, and application of information are core problems for many of a university’s most central endeavors. Because of this, I School perspectives and inputs can enhance university administration and management, expand the reach and value of scholarly communications, and open new horizons for far more accessible and effective instruction. Despite this promise, there are several important challenges that must be faced before we realize the potential of fostering more significant and effective partnerships across different parts of the university community.

Our roundtable discussion will explore strategies that can facilitate more effective university-wide (and beyond) partnerships. Workshop participants will share ideas about how to foster collaborations between I Schools and other campus groups to better address challenges and opportunities faced across the scholarly community. Among the topics to be explored will be:

1. Marketing I Schools and I School perspectives
2. Prioritizing projects, and aligning interests to form collaborations
3. Communicating across disciplinary divides
4. Identifying cross-cutting solutions that may work in several different application areas

Part of this discussion will draw from experiences gained through an experimental program now underway at the UC Berkeley School of Information. The Information and Service Design (ISD) program, launched in January of 2007, is part of the university-wide initiative in Service Science, Management & Engineering and emphasizes the unique people and perspectives of the UC Berkeley School of Information. The ISD Program explores increasingly global and interconnected developments in business, law, computing, communications, research, and education.

This program runs an "Information Systems Clinic" which is open to students from engineering, computer science, business and other disciplines. Students benefit from this program by having numerous opportunities to work with faculty on service-related research, as well as hands-on design and implementation projects with both on-campus and off-campus organizations.

This Clinic helps facilitate and coordinate projects with units across the UC Berkeley Campus. It now includes a portfolio of some five related projects intended to enhance open access, usability, and impact of essential scholarly resources. These projects are developed in collaboration with UC Berkeley Libraries, museums, "cyberinfrastructure" (scientific communications) initiatives, as well as leading nonprofit organizations including the Internet Archive and Creative Commons.

ROUNDTABLE SUBMISSION 3:

Science and Technology Studies (STS) in iSchools

AUTHOR(S):

Van House, Nancy
Blanchette, Jean-Francois
Borgman, Christine
Lievrouw, Leah
Star, Susan Leigh

ORGANIZATION(S):

UC Berkeley, United States of America;
University of California, Los Angeles, United States of America
Santa Clara University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The interests of Science and Technology Studies (STS) and iSchools have converged in recent years. Building upon a successful five-session track on the influence of IS on STS at the recent Social Studies of Science Conference (October, 2007, Montreal), we propose a complementary session at the iConference that reflects on the influence of STS on IS.

STS is an interdisciplinary field that studies technology, the mutual constitution of technology and social, and sociotechnical networks; the practice of science and the development of scientific knowledge (and, by extension, other kinds of knowledge and knowledge communities); and national and international science policy. STS generally takes a strongly constructionist, situated, social, and critical perspective on these topics. Both fields are concerned with the interaction among information and communication technologies (ICTs), knowledge, society, practice, the social, and public policy. And both are highly interdisciplinary: STS' roots are in history, philosophy, sociology, and anthropology.

STS is increasingly represented in iSchools. Many STS scholars are finding congenial homes on iSchool faculties. Many iSchool faculty and PhD students are doing research drawing on STS concepts and methods. STS courses are being offered in iSchools, and iSchool students are taking STS courses in other departments.

These shared interests are apparent in the world of scholarly publishing and conferences. Many iSchool faculty and PhD students have presented at the Society for Social Studies of Science (4S) annual conference in recent years. Some publications have looked at STS and IS, notably Van House's 2003 ARIST chapter "Science and Technology Studies and Information Studies," and, most recently, Boczowski and Lievrouw's "Bridging STS and Communication Studies: Scholarship on Media and Information Technologies" in the just-published third edition of the Handbook of Science, Technology and Society. Borgman's new book, *Scholarship in the Digital Age*, draws heavily on both STS and IS theory and practice. Bowker and Star's *Sorting Things Out: Classification and Its*

Consequences have been highly influential in both fields, and Bowker's Memory Practices in the Sciences won "best book" awards from both ASIST and 4S, while Star is the immediate past president of 4S.

We are proposing a panel to examine this growing alliance. We will briefly explain to iSchool people who are unfamiliar with STS what it is and why it's increasingly intertwined with iSchools. We will also propose to the panel a series of questions about the pros and cons of this alliance, and possible future directions for research and teaching. These questions may include:

- In what areas of research and teaching can STS contribute to iSchools' concerns? And vice versa?
- Where are the ideological/discursive/theoretical overlaps and disjunctions between STS and iSchools?
- What are STS methods how are they usefully included in iSchools? In research? In the curriculum?
- Does STS have a place in the master's curriculum?
- How does STS help build alliances on campus?

ROUNDTABLE SUBMISSION 4:

Training, Integration, and Identity: A Roundtable Discussion of Undergraduate and Professional Master's Programs in iSchools

AUTHOR(S):

Hemphill, Libby

Munson, Sean A.

ORGANIZATION(S):

University of Michigan, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Professional students, whether undergraduates or masters' students, represent a significant portion of the iSchool community. How do iSchools effectively educate those students while continuing to develop successful research programs? This roundtable discussion will focus on how iSchools educate their professional students and engage them in the research aspect of their programs. Innovative approaches to training and integration will be the central theme of this discussion. In an iSchool – where students train for professions including librarianship, information policy, human-centered computing, preservation and researchers explore such topics as incentive-centered design, forensic informatics, computational linguistics, and digital libraries have both competing and complimentary goals – the potentials for collaboration, innovation, misunderstanding, and disharmony are all high.

The annual iConference provides a unique opportunity for us, as a community, to discuss the roles our professional students have in shaping our identity and our practices. The proposed roundtable will invite participants to discuss questions such as:

- What should the role of research in training information professionals be?
- How can we best engage professional students in our research?
- How do iSchools address the unique curricular challenges we face in preparing students for a very wide variety of careers?
- What do we want an Information degree to signal in the marketplace?
- What are some successes in which research and professional training have benefited one another?

Participants will share innovative approaches to professional education, best practices in engaging professional students in research programs, and remaining challenges. We intend roundtable participation to represent the diversity of iSchools' current programs.

Confirmed participants include:

- Dr. Eileen G. Abels, Master's Program Director, Associate Professor, College of Information Science & Technology, Drexel University
- Dr. Judith S. Olson, Richard W. Pew Collegiate Professor of Human Computer Interaction and Associate Dean for Academic Affairs, School of Information, University of Michigan

Each speaker will present introductory remarks highlighting some of the achievements

and challenges they face in their home programs, after which discussion will include questions and input from the attendees. This will be an interactive forum proposing ideas for new approaches to education and integration of professional students. We encourage participants to discuss ideas that work (and those that don't!) in their schools. We will create and publicize a wiki space for pre- and post-conference participation as well.

ROUNDTABLE SUBMISSION 5:

Information: Transforming the World through Better Communications

AUTHOR(S):

Dietel, Ronald James
Welshons, Marlo
Shafer, Kelley
DuBois, Charles

ORGANIZATION(S):

University of California, Los Angeles, United States of America
University of Illinois at Urbana-Champaign, United States of America
University of Pittsburgh, United States of America
Pennsylvania State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Introduction: A central tenet of the iSchools is that “the right information delivered at the right time to the right people in the right form,” will “provide society with the key to success.” (www.ischools.org/oc/study.html) For this goal to be realized, information scholars and professionals must effectively communicate their knowledge to those people who will do something useful with it. The purpose of this roundtable is to increase the communication skills of session attendees in support of both iSchool and iConference 2008 goals. More specifically, this communications roundtable will provide time-tested communications tips to improve participant’s ability to focus their information message, identify key audiences, and improve their presentation, writing, and media skills. The session will also draw on the diverse positive and negative experiences from attendees, sharing their personal experiences in communicating effectively with different audiences and through different formats.

Audience: We expect that the audience for this communications roundtable will be multi-cultural and represent multi-disciplinary information sectors, thus contributing to the value of the session and the Call for Participation. We believe that this roundtable will serve an important mentoring function as stated in the Call and simultaneously support the Doctoral Colloquium program.

Content: This communications roundtable will begin with a solicitation from participants of both good and bad personal experiences they have experienced, which might range from a conference paper being read to them, to a misquote in a newspaper. Such communications mishaps are far too common, but the roundtable leader will share with attendees a method that traditionally helps improve communications skills. These include: 1) belief in the value of improved communications, 2) clarifying the purpose of the communication, 3) audience identification, 4) selection of appropriate media, 5) message development, 6) message delivery, and 7) evaluation of results. The session will also describe the goals of the iSchools communicators, including the development of the iSchools logo and web site.

A majority of time will be spent on message development, with session leaders providing a list of top ten suggestions for 1) improving writing skills, 2) making presentations, and 3) working with the media. Attendees will be provided with a written list of tips in each major topic, plus a reference list of resources for communications improvement. We recommend a minimum of 90 minutes to cover the proposed content and to ensure useful participant involvement.

Evaluation: Attendees will be asked to complete an evaluation form for the session, with a goal of improving similar sessions at future iConferences.

About the Presenter/s: Supporting the proposed objectives, a group of three veteran university communicators will share more than 30 years of experience helping researchers and other information specialists formulate and communicate their work to diverse audiences.

Ronald Dietel, UCLA Graduate School of Education & Information Studies

Marlo Welshons, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign,

Kelly Shafer, School of Information Sciences, University of Pittsburgh

ROUNDTABLE SUBMISSION 6:

Social Network Analysis

AUTHOR(S):

Yen, John

Giles, Lee

Foley, Hank

Adamic, Lada

Menczer, Filippo

Wang, Haizheng

Tapia, Andrea

ORGANIZATION(S):

Pennsylvania State University, United States of America

University of Michigan, United States of America

Indiana University, United States of America

Microsoft Corporation, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The Web transformed internet to a global information sharing infrastructure; but the online social network websites further transformed the Web into a global cyber community. As a result, people perceive the Web increasingly as a social medium that fosters interaction among people, activities of groups, formation and evolution of communities, for sharing information, knowledge, experiences, and even wisdom.

The social flair of the Web poses new challenges and opportunities that iSchools are uniquely positioned to address. Social networking in the Web is a phenomenon of scientific interest from multiple disciplines; there is demand for interdisciplinary research on the rules and behavioral patterns that emerge and characterize community formation and evolution. Further, the social Web introduces new behaviors for an on-line community: searching for similar people, searching for mentor/mentees, and searching for answers by combining blogs/answers from multiple people. Moreover, the misuse potential of social formations cannot be stressed enough: People are confronted with unreliable or even malicious content, with surveillance and even stealth of information and of personal data; detection and protection mechanisms are needed here, based on a deep understanding of patterns of misuse and misbehavior. Finally, the social structures in the Web is very diverse: including wikis and folksonomies where people contribute semantically rich content, platforms for collaborative annotations where people enrich existing content, bulletin boards and similar fora where people seek for advice but also establish new contacts.

Researchers of iSchool are well positioned to address these challenges by integrating perspectives from sociology, information science, scientometrics, network science, and computational informatics. The integration of these conceptual frameworks is likely to not only offer new insights for the modeling and the analysis of social networks, but also to post new questions that are now made possible by the availability of the huge volume of data regarding a wide range of networks. For instance, the study of collaborator networks can shed lights on the relationship between their growth and the popularity of the ideas central to the network; which can be used to predict the growth of network and/or the epidemiology of ideas.

The proposed roundtable will bring together researchers from the iSchool community to discuss trends, new questions, and innovative ideas regarding social networks. For instance, how to discover and analyze subcommunities within a very large social network? How new behaviors in on-line social networks emerge through interactions? How social networks hidden within blogs, fora, and other forms of on-line communications can be identified? What are the impacts of social networks on social activities? How do social networks form, evolve, and grow? How ideas are spread over time and space using social networks? The proposed roundtable will stress the interdisciplinary challenges of these and other topics related to social networks.

Through these interactions, we hope the roundtable can help the iSchool community in establishing one or multiple grand challenges regarding social networks and launching an effort to develop a strategy that aims to position iSchools to be leaders in addressing these grand challenges.

ROUNDTABLE SUBMISSION 7:

Whose Turn is It? Research on Gaming in the iSchools

AUTHOR(S):

Nicholson, Scott
MacInnes, Ian

ORGANIZATION(S):

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

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 can not provide. Libraries are supporting gaming activities and educators are integrating gaming in new ways. Most of the younger generations (and many of the older generations) 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 caught 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 teaches 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 several 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. A goal will be to craft several potential multi-disciplinary and multi-institutional grant proposals in order to allow the iSchools to take their turn at the gaming table.

ROUNDTABLE SUBMISSION 8:

Young People and Technologies: Fostering Transformative Experiences

AUTHOR(S):

Tilley, Carol, L.

ORGANIZATION(S):

University of Illinois at Urbana-Champaign, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

In "Preparing School Library Media Specialists for the New Century: Results of a Survey" (Journal of Education for Library and Information Science 42: 3, pp. 220-227, Summer 2001), Carol Tilley and Daniel Callison found that among schools accredited by the American Library Association technology-focused courses ranked highest on the list of the most widely required courses for this professional speciality. The survey also revealed that technology-related courses dominated the roster of elective coursework.

A quick reading of the survey may suggest that these graduate programs in information studies had presciently understood the increasing role that information and communications technologies (ICTs) play in the daily lives of both young people and the information professionals who serve them. Yet, data from the survey also revealed that ICT-related coursework focused on ICTs in service of professionals' needs, not ICTs in service of youth empowerment. Furthermore, the survey's scope did not allow it to address more illuminating questions including the extent to which other youth services information professionals such as public library children's specialists receive training in ICTs, to what degree education related to ICTs is supplanting a focus on traditional media and technologies, or how information schools can prepare professionals to foster transformative experiences for young people through the use of ICTs.

The purpose of this roundtable, then, is to provide a forum for discussing how information schools might more effectively educate youth services information professionals in the theory and application of ICTs to their interactions--structured and unstructured--with young people. Participants will be encouraged to bring relevant course descriptions, class syllabi, assigned readings, and course assignment description to the discussion to provide concrete examples of issues. The conversation will be enriched through references to appropriate models from community and social informatics, media literacy, and traditional librarianship, as well as research and best practices in education. The Pacific Bell/UCLA Initiative for 21st Century Literacies provides an additional corpus of examples, research, and practice on which to draw.

ROUNDTABLE SUBMISSION 9:

The Role of iSchools in Medical Informatics

AUTHOR(S):

Haque, Saira Naim
Oesterlund, Carsten Svarr
Reddy, Madhu C

ORGANIZATION(S):

Syracuse University, United States of America
Pennsylvania State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Similar to iSchools, healthcare is an interdisciplinary field that encompasses heterogeneous environments (e.g., hospitals, clinics, home) and perspectives. However, medical informatics research focuses mainly on technical solutions to narrow problems. Often, this leads to the design of systems that solve one problem but do not fit the general work practices in a setting. In response, there is a growing movement in the community to start investigating organizational and social issues that impact the design, implementation, and use of information technologies (Reddy and Bradner 2005; Kaplan and Shaw 2004). This trend will only increase as healthcare documentation in the US increasingly moves from paper-based to electronic systems.

iSchools are uniquely positioned to benefit the medical informatics community due to our interdisciplinary foundation. Taking the theoretical paradigms and methods from various fields will allow us to address the interplay between people, healthcare technologies, and healthcare organizations. Some areas could include:

- Design and evaluation mechanisms to account for the ways different stakeholders interact with systems.
- Workflow analysis to understand how information is used for various activities in preparation for a systems implementation.
- Developing and implementing systems that deliver the right information to the right people at the right time.

iSchool research can also help expand the focus of the medical informatics community to examine other important issues such as the digital divide in healthcare and protecting patient privacy. By working together, the iSchools can train the next generation of researchers providing needed interdisciplinary insight into medical informatics issues and extend into new territories. Healthcare is characterized by multiple stakeholders and teams that are separated by time and space that are information-dependent; therefore, iSchool research in this field can also impact research in related areas. Besides being a rich setting for conducting research, healthcare impacts our lives on a daily basis.

The hosts of the roundtable will discuss how iSchools can get involved to incorporate more interdisciplinary research in the medical informatics area. In particular, we will focus on the following questions:

1. What are the interesting medical informatics research problems from an iSchool perspective?
2. How can iSchool research impact the medical informatics field?
3. What are the challenges iSchools face in this field?

Hosts

Saira Haque (<http://snhaque.wordpress.com/>) has held a variety of healthcare management positions, including consulting and hospital administration. She is the doctoral student member at large of the American Medical Informatics Association and a Diplomat of the American College of Healthcare Executives. Her research focuses on coordination of interdisciplinary work across healthcare settings.

Carsten Østerlund (<http://blogs.iis.syr.edu/osterlund/>) has studied documenting work in healthcare settings over the past decade. He explores how the materiality of different document genres (e.g., paper, electronic, wall mounted) interact with the distributed collaboration among various professional communities.

Madhu Reddy (<http://faculty.ist.psu.edu/reddy/>) has been involved in the medical informatics community for a number of years. His research has focused on collaborative technologies in healthcare. He is currently the co-chair of the Organizational and Social Issues Working Group, International Medical Informatics Association.

References

- Reddy, M. and E. Bradner (2005). Multiple Perspectives: Evaluating Healthcare Information Systems in Collaborative Environments. *Evaluating Healthcare Information Systems: methods and applications* 2nd ed. J. Anderson and C. Aydin. New York, Springer.
- Kaplan, B. and N. Shaw (2004). Future directions in evaluation research: people, organizational and social issues. *Methods of Information in medicine*. 43(3): p. 215 - 31.

ROUNDTABLE SUBMISSION 10:

Mediating Intimacy: Navigating friendships, family ties, and romantic relationships with communication technologies

AUTHOR(S):

Cheshire, Coye
Fiore, Andrew T
Sims, Christo

ORGANIZATION(S):

University of California, Berkeley, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Social relationships with friends, family, and romantic partners provide structure and meaning in our lives. We continually gather information through interpersonal interaction, which influences our motivations and social behaviors. With the widespread adoption of technologies that mediate interpersonal communication, many of these longstanding practices are being reconfigured. Many mediated social practices have no

analogue in face-to-face interaction; others violate norms and expectations in subtle ways, such that our social intuitions, honed offline, may lead us astray.

Despite the adaptations that mediating technologies such as mobile phones, email, and text messaging require, our goals in interpersonal interaction remain the same: identifying and connecting with potential friends and romantic partners, maintaining existing close relationships, and sometimes dissolving those relationships when they sour. Different mediating technologies both constrain and facilitate these processes to different degrees. Online dating systems, for example, facilitate finding a romantic partner by providing detailed information about dozens or hundreds of potential mates; however, nuances of personality and behavior are difficult to discern within the constraints of asynchronous, text-based messaging that many systems provide. Similarly, text messages on a mobile phone facilitate quick, non-intrusive coordination and signaling of presence, but they are too short and the typical input method too cumbersome for in-depth discussions (though the teenage text-messaging patois works to alleviate this).

We propose a roundtable discussion of current theoretical and empirical work as well as future research directions relating to the management of intimate relationships through mediating technologies. The discussion will be open to all conference attendees, and furthermore we will try to identify interested faculty and students ahead of time. Several graduate student researchers at the UC-Berkeley iSchool (including those who collaborated to write this proposal) have already expressed interest in participating in this roundtable session.

We would like to explore the following topics in the discussion:

- How do people take advantage of mediating technologies to foster intimacy in new ways? In situations not traditionally structured for intimacy?
- How do people circumvent the constraints of these technologies to recover some of the intimacy they feel face-to-face?

Roundtable Proposal: Mediating Intimacy

2

- How do mediating technologies change the nature of public and private space and time among intimate friends, partners, or family?
- How might we design communication technologies that better support the formation and development of intimacy?
- Given the possibility of continuous contact by technological means, how do individuals negotiate for and create space apart from their intimate others? How do people manage the expectations and obligations that come with constant availability?

ROUNDTABLE SUBMISSION 11:

The Globalization of Education for Digital Librarianship –Implications for iSchools in North America.

AUTHOR(S):

Weech, Terry
Moulaison, Heather

ORGANIZATION(S):

University of Illinois, United States of America
Rutgers University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

In the United States, members of the iSchool group have received grants to develop education programs for information professionals to work with digital library collections, but have taken distinctly different approaches on how to package and offer their educational programs. Some of the schools have focused on independent certificate programs and still others on Master's degree programs and post-Master's degree programs.

In the rest of the world some very different approaches have been taken to providing education for digital librarianship. There are two distinct models in European education for digital librarianship. The most common approach is to follow the model of integration of education for digital librarianship into existing information studies

programs. A second model follows the North American iSchool approach of maintaining separate programs for digital librarianship. There are only one or two programs in Europe that follow this separate program approach. One innovative example of the alternative program approach is the International Master's in Digital Library Learning (DILL) which is a European Union funded consortium of three European Information Studies schools. This consortium provides specialized education on digital libraries at the "post-graduate" level in a two year Master's Degree program with students recruited internationally. Students from North America are included in their recruitment program, as well as students from Asia, Africa, and South America. The program consists of four course modules offered over the two year DILL program. Currently a partnership with a North American iSchool is being explored to further the international aspects of the course of study.

The proposed roundtable discussion will invite representatives from iSchools in North America to discuss their digital library education programs and any international connections of their programs. One or more representatives from an international program of education for digital librarianship will summarize the status and nature of international planning for digital librarianship education. The roundtable discussion will be organized and led by Terry Weech, Associate Professor at the Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign with the assistance of Heather Moulaison, Rutgers University's School of Communication, Information and Library Studies. Terry Weech has chaired the IFLA Standing Committee on Education and Training and conducted an international study of education for digital librarianship. He is currently involved in an international study of equivalency and reciprocity of qualifications for graduates of information study programs and is working with the International Master's in Digital Library Learning (DILL) consortium to explore the possibility of expanding the consortium to North America. He also participated in the Morocco Digital Library Workshop on "Implementing a Maghreb Digital Library for Education, Science & Culture" Rabat, Morocco, 25-29 January, 2007, which was sponsored by the US National Science Foundation, UNESCO, and the Fulbright Academy of Science and Technology. Heather Moulaison has helped Moroccan contacts investigate the possibility of digital libraries as a knowledge management tool.

ROUNDTABLE SUBMISSION 12:

Bridging the Divide Between Theory and Practice in ICT for Development

AUTHOR(S):

Ho, Melissa Rita

Veeraraghavan, Rajesh

ORGANIZATION(S):

University of California, Berkeley, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Recent years have seen a burgeoning interest in the use of information and communications technologies for development (ICT4D/ICTD). In academia, this has manifested itself as evaluations of Internet kiosks (e.g. e-Choupal, Akshaya, telecentre.org) and other ICTD interventions, as well as projects like One Laptop Per Child (OLPC), Wireless Long Distance Networks (WiLDnet) and MultiMouse, which endeavor to innovate new technologies that address the specific needs and requirements of communities in developing countries. We propose to hold a roundtable discussion, in which PhD students interested in or currently working on questions around information and communications technology in emerging regions can discuss some of the overarching issues entailed in our research topics. In particular, we examine the multi-disciplinary nature of our research, and the role of I School students in understanding where theory and practice meet in ICTD.

Information schools, as a place where social science and technological practice intersect, are well positioned to do this research. Technologists often look at ICTD from an engineering and implementation perspective. Social scientists have a good understanding of the social structures and other factors underlying ICTD deployments. However, communication between the two groups is often limited by a lack of shared vocabulary, and a difference in practice. As I School students, we take an interdisciplinary view on this question, which we hope will result in a more synthesized and well-grounded approach to ICTD interventions.

As with other interdisciplinary endeavors, our challenge is to bridge the divide between theory and practice, understanding how our efforts in understanding ICTs and implementing new ones can contribute to academia. What is the role of social theory and development theory in real-world ICTD deployments? How can existing bodies of knowledge contribute to ICTD practices? Can there be a theoretical basis for practical deployments? We

believe that the answer is an affirmative “yes,” but there is no clear and obvious path for us to take in carrying out theory-driven ICTD research projects. For that matter, there are no clear and obvious venues in which we should publish the results of our research.

We have two main goals for this roundtable discussion:

1. Community Building: Identify and share experiences with students from other I Schools also interested in this topic, discussing ways in which we can better support each other across schools.
2. Approach: Discuss the challenges of doing ICTD research, specifically addressing the challenge of linking theory and practice in the work that we do.

ICTD work necessarily ranges across a wide variety of disciplines, encompassing (but not limited to) fields like public health, development policy, business, education, development theory, human-computer interaction, and networked systems. As individuals, we are not able to be experts in all of these areas simultaneously. As a community, however, we will be able to leverage one another's expertise, building a collective understanding of the work we endeavor to carry out.

ROUNDTABLE SUBMISSION 13:

Information Assurance, Intelligence and Security: Opportunities and Directions for Future Research

AUTHOR(S):

Paquette, Scott

Fagnot, Isabelle

ORGANIZATION(S):

University of Maryland, United States of America

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

A growing number of the iSchools have recently begun to emphasize programs of research directed towards the topic of information assurance, intelligence and security. Covering such areas as network and infrastructure security, information systems auditing, behavioral security, information and technology risk management, and business continuity planning, the idea of information security has become a focus of not only iSchools but major corporations and intelligence agencies. This field is also receiving much attention in academic institutions as new courses, programs, and research centers are being created.

This roundtable would present a picture of the current state of research in the information assurance and security field including current trends, directions and new areas of emphasis for future research, how the iSchools have the potential to develop unique contributions to the field, and possible collaborations between iSchool researchers and other departments. Hosted by a combination of tenure-track professors and PhD candidates with research and/or professional experience in the information assurance, intelligence and security field, this roundtable would introduce attendees to significant research opportunities and the role of iSchools in this growing field.

The roundtable will be organized with brief presentations by the panelists who will each present a perspective on this area based on their previous experience and research, followed by a discussion with attendees. The first panelist, Scott Paquette of the University of Maryland will discuss how security can be viewed from an organizational perspective and combined with research on assurance and risk management. Isabelle Fagnot from Syracuse University will examine information security and assurance from a more individual and behaviour perspective, focusing on research that combines information science and other social disciplines. Kevin Desouza from the University of Washington will present issues related to security for the safe guarding of knowledge assets and intellectual capital, especially as it relates to the complexities found in intelligence agencies. Dave Yates of the University of Maryland will examine issues related to copyright and privacy, and how security at all organizational levels can contribute to these issues. Finally, Shuyuan Mary Ho from Syracuse University will present some paradigm shifts from the history of information assurance and research gaps identified through analyses of intrusion. This diverse panel will emphasize the vast and varying topics within information security and assurance, differing levels of analysis, implications of this research on the iSchools and other fields. As well, they will demonstrate that the concepts of information assurance and security can be integrated into many current research agendas. After the introductory presentations, Scott Paquette will moderate a discussion between the panelists and roundtable attendees.

The goal of this roundtable is to present different views on the state of assurance, intelligence and security research, and to create a forum where new ideas and opportunities can be illuminated and discussed by all participants. Conference attendees new to and already familiar with this topic should take away an understanding of the importance of this field, the potential in research opportunities, and how they can contribute to its growth based on their own research strengths and knowledge.

ROUNDTABLE SUBMISSION 14:

Intersectionality and Interdisciplinarity: Information Studies and Studies of the “Other”

AUTHOR(S):

Chu, Clara M.

ORGANIZATION(S):

University of California, Los Angeles, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Although long under-emphasized in information research and education, information studies (IS) has an increasing interest in the “Other.” By Others we refer to those groups that are subordinated or excluded from dominant society or culture(s), and have been organized into particular categories of race, ethnicity, gender, sexuality, class and development, among others. This roundtable discussion aims to examine the ways in which information studies and other fields that study the Other, such as ethnic studies, queer studies, women’s studies, area studies, among others, engage in intellectual discourse, scholarship and education to enhance research and learning. The World Café conversational process will be used to guide the discussion. “The World Café is an innovative yet simple methodology for hosting conversations about questions that matter. These conversations link and build on each other as people move between groups, cross-pollinate ideas, and discover new insights into the questions or issues that are most important in their life, work, or community.” (Source: <http://www.theworldcafe.com/>) The two main questions, which frame the discussion are: (1) have points of intersectionality been explored and identified to address issues of equity of information access? and (2) has information studies developed an interdisciplinary relationship with any of these fields?

For the purpose of the discussion we will use the following working definitions: ““Intersectionality” refers to the fact that human identity is indivisible and that the struggle against one form of oppression (such as heterosexism/ homophobia/ lesbophobia/ biphobia/ transphobia) cannot in practice be separated from the many other struggles that members of our communities are engaged in (such as struggles against sexism, ableism, racism, or economic disadvantage). To address oppression effectively, it must be addressed holistically.” (Egale Intersections Committee, 2002 <http://www.egale.ca/index.asp?lang=E&menu=40&item=302>)

“Multidisciplinarity is a recognition of disciplines working together without integrating their professional or scholarly assumptions or epistemologies. In contrast, “interdisciplinarity” will be used to refer to two or more disciplines that do develop a shared research or subject content interest and integrate to some degree their methodological and conceptual view of the other disciplines in their shared efforts.” (Weech, 2007)

Related questions that will be explored are: Which relationships exist and how did they originate? Are they uni- or bi-directional? What research or programmatic initiatives exist? How are they funded? Which iSchools have faculty, programmatic initiatives or mandated curriculum that enable interdisciplinarity between the above fields? The session will conclude by identifying the benefits and challenges of interdisciplinarity and intersectional exploration between information studies and studies of the “Other” for their respective scholarship and education and in addressing the “isms” in society.

Wildcards

(17 Accepted Submissions)

WILDCARD SUBMISSION 1:

Mapping the Design Space of Design Education in iSchools

AUTHOR(S):

Jones, M. Cameron
Floyd, Ingbert R.
Twidale, Michael B.
Adamczyk, Piotr D.

ORGANIZATION(S):

University of Illinois at Urbana-Champaign, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Mapping the Design Space of Design Education in iSchools

M. Cameron Jones, Ingbert R. Floyd, Michael B. Twidale, Piotr D. Adamczyk

University of Illinois at Urbana-Champaign

iSchools have a well-established reputation for rigorous analysis, applying a range of multidisciplinary methods. However, increasingly many of us are teaching design elements as well as doing design in our research. We believe that as iSchools, we have a unique perspective, or accent, on design activities, which comes from our history of studying not just the design of information systems and how people use them, but from taking a step back and studying the information behavior of individuals and groups, thus taking a more holistic view of design. It is time for iSchools to assert their accent more explicitly in the study of design by integrating design perspectives across multiple disciplines. The development of new multidisciplinary design schools, such as the Stanford Design School, and RPI's Product Design & Innovation program, highlight the value of an iSchool-like, multidisciplinary approach to understanding design. However, iSchools' accent is more than just a multidisciplinary approach; it can bring a critical perspective on multidisciplinary design that can also question the function of design as an information practice. iSchools can account for ethical, social, and more systematic concerns, along with an accent on the more overlapping cognitive and information aesthetic aspects of design practice. However, while this new accent on design is present in iSchools, it is currently underarticulated.

With the recent growth of computer applications that can be tailored and combined without requiring sophisticated programming skills, and with the greater accessibility to tools for technology production, design is becoming an increasingly essential skill in the modern organization. Students need to learn design skills and "design-thinking" in order to secure desirable jobs post-graduation [2]. What are we, as iSchools, doing to prepare students for this emerging market? How are we teaching design, design practice, and design thinking? In art and architecture programs, students take design studios where they learn design methods and engage in long-term design projects. However, in iSchools and Computer Science programs students are typically given little explicit design instruction, and are often forced to discover on their own how to think about and do design as they engage in projects for class and work.

We are proposing this Wildcard to enable a conversation about design from an iSchool perspective. We aim to focus on the special case of design instruction. The proposal authors have experience teaching design at different levels (undergraduate, masters, and doctoral) to multidisciplinary audiences, with students not just from more systems-construction-oriented disciplines such as CS, Library and Information Science (LIS), Urban Planning, Electrical and Computer Engineering, and Geographic Information Systems; but also from disciplines as disparate as English, History, Media Studies, International Studies, Journalism, Cognitive Science, and various disciplines within Education (Curriculum and Instruction, Secondary and Continuing Education, Educational Psychology, and Human Resource Education) [1, 3]. For many of our students, however, thinking in design terms is an alien concept. Thus, we see this Wildcard as having the secondary goal of helping us all to understand the design space of teaching design to students with diverse backgrounds. The aim of this Wildcard is to consider a particular design space that of teaching design in an iSchool. We aim to chart our collective experiences and intentions in teaching to help understand the challenges of conflicting goals, various opportunities and multiply-scaled constraints that characterize this particular design

space.

As designers ourselves, we are aware of the dangers of locking in on particular ideas too early in the process. Thus, we envision this Wildcard as being a brainstorming session for initializing this discussion of what we each mean when we talk about design. To accomplish this work, we intend to utilize the very methods which we teach for exploring design spaces. At the Eclectic Design Research Group (<http://design.lis.uiuc.edu/>), our motivating philosophy is to create as many designs as possible, as quickly and cheaply as possible, and then iteratively explore the design space to test them out, evaluate them, integrate them, and improve them. We see this session as a brainstorming activity where we will attempt to apply this philosophy to the design of our classroom experiences. An incomplete list of possible brainstorming techniques we will be using includes: Concept Mapping, Bad Ideas [4], Scenarios, and Personas.

There are various questions that we believe this activity will allow us to explore. We do not expect to cover all of them exhaustively, and to a large extent will leave it open to the participants which ones they want to concentrate on. Participants will be able to identify particular questions of interest via the electronic infrastructure we set up.

Some of these questions are:

- What is design? Where does design intersect iSchool curricula?
- What are the methods of design which are meaningful and useful in iSchool education?
- What does a class devoted to design look like?
- Who participates in design? Who can do design?
- Where does design happen? Where doesn't design happen (and it could)?
- What are the particular design concepts we are trying to teach?
- What are barriers to teaching design? How can we overcome those obstacles?
- Is there an iSchool way of teaching design? How is that different from other disciplines?
- Are we (or should we be) teaching design or design thinking? Is there a difference?
- How does the iSchool perspective on design compare with the growing discussion of Computational Thinking [5]?
- How do we teach design across many contexts, to students from many backgrounds?
- What are the contexts of design in iSchools? Where is design currently applied? Where is it not?

We hope that this Wildcard will be the initiation of an ongoing, persistent conversation around these topics. Prior to the conference date, we will be setting up electronic infrastructure (<http://design.lis.uiuc.edu/ed/>) for conducting preconference planning activities. We are not sure exactly what form this will take, but we are considering the use of wiki, blog, mailing-list, and document-sharing technologies, possibly supplemented with social networking services. We will be inviting potential participants to share in the planning activities of the Wildcard, by contributing suggestions and topics for discussion, and by helping us create seed materials for facilitating the collaborative, face-to-face brainstorming activities we plan to conduct at the conference.

Depending on the number of participants, we may invite them to present a two-minute summary of one teaching experience in order to motivate and contextualize the discussion. These

presentations might be an account of what was tried, what the motivations and inspirations were, or how the presenter went about trying it. It might be a success, a failure, or an effort that just went, "meh". We are looking for both the tried-and-true, as well as cutting-edge pedagogical techniques. Informal assessments, qualitative analyses, quantified measures, and any other kinds of evaluations are all welcomed and encouraged so that we can explore the broadest design space possible. The emphasis of the session is not on finding answers, but on finding questions, possibilities, and productive venues of exploration. Even preliminary answers will have to wait until next year.

The actual Wildcard session will consist of the following:

Time	Activity
0-15 Minutes	Introduction to the session, where we describe the purpose of the session, and if we don't have too many participants, do quick introductions.
15-60 Minutes	Break into small groups for brainstorming sessions, each of which will utilize a method of brainstorming design. The activities will involve the sharing of particular design-teaching experiences to help map out the space. For example, one might be a listing of different misconceptions students may have about design that instructors try to address, and another might be different activities and interventions used to address a given misconception. Time permitting, we'll break up and reform into different groups and iterate the brainstorming session.
15-90 Minutes	Debriefing where we discuss what we brainstormed, and prepare to continue work after the conference via the electronic infrastructure.

Table 1. Schedule of Wildcard Session Activities

The goal of this session will be to begin to map out the different aspects of the design space of teaching design in order to characterize the design space as fully as possible. Following the conference, we will be continuing the articulation work we began at the conference via the electronic infrastructure. Participants, and others interested in design education, will be able to take topics and ideas generated in the session and explore them further in their classroom and research activities. Results from these explorations can be reported back, updating our collective understanding of design and design education, allowing us to refine the evolving map of the design space. We also anticipate an ensuing discussion of the design of the evaluation of teaching design -- how do we determine the success of not only the students' ability to learn and do design, but also the instructor's ability to effectively teach design?

At next year's iSchool Conference, we will reconvene in a face-to-face session, and participants will present on their experiences over the preceding year. New participants will be encouraged to join and contribute their experiences as well. We will use the remainder of the session to reevaluate where we stand and reflect on what we've learned about design and design education, as well as map out new directions and opportunities for improving design instruction in iSchools. Thus, a part of this exercise will be exploring conferences coupled with persistent conversation mechanisms as a venue for both initializing research and for advancing the field.

References

- [1] Adamczyk, P.D., M.B. Twidale (2007) Supporting Multidisciplinary Collaboration: Requirements from Novel HCI Education CHI 2007.
- [2] Brown, T. (2005). Strategy by Design. Fast Company, 95, p. 52, June 2005.
- [3] Dix, A., Ormerod, T., Twidale, M., Sas, C., Gomes da Silva, P., McKnight, L. (2006). Why bad ideas are a good idea. In Proceedings of HCIEd.2006-1 inventivity, Ballina/Killaloe, Ireland. 23-24 March 2006.
- [4] Floyd, I. R., Jones, M. C., Rathi, D., Twidale, M. B. (2007). Web Mash-ups and Patchwork Prototyping: User-driven technological innovation with Web 2.0 and Open Source Software. To appear in Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07). Waikoloa, Hawaii, USA.
- [5] Wing, J.M. (2005) Computational Thinking CACM, 49(3) 33-35.

WILDCARD SUBMISSION 2:

Is there a cloud in your future? Applications of “cloud computing” to Web-scale problems

AUTHOR(S):

Lin, Jimmy

ORGANIZATION(S):

University of Maryland, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Is there a cloud in your future?

Applications of “cloud computing” to Web-scale problems

Proposal for a “wildcard” session, iConference 2008

Organizer:

Jimmy Lin (jimmylin@umd.edu)
Assistant Professor
College of Information Studies
University of Maryland, College Park

1. Background

IBM and Google recently committed a total of \$30 million over two years to an initiative on “cloud computing”, in collaboration with six universities across the country (see references). They are: Berkeley, Carnegie Mellon, MIT, Stanford, the University of Maryland, and the University of Washington. I am the leader of this initiative at the University of Maryland, and to my knowledge the only participant from an iSchool (the rest are lead by faculty in computer science departments).

“Cloud computing” refers to technology for exploiting large computer clusters to tackle “Web-scale” information processing problems, where immense quantities of data make traditional sequential processing impractical. Specifically, this initiative focuses on Google’s MapReduce programming paradigm, which was specifically designed for processing extremely large data sets (and indeed used by Google itself for much of its production operations). Programs written in the MapReduce functional style are automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program’s execution across a set of machines, handling machine failures, and managing the required inter-machine communication. Hadoop is an open-source implementation of the MapReduce framework.

As a part of this initiative, IBM and Google are making Hadoop clusters available to the university collaborators, with the simultaneous goal of advancing research and education. For the past two months, the Computational Linguistics and Information Processing (CLIP) Lab at the University of Maryland has been actively exploiting this resource for research in natural language processing and information retrieval.

The exponential explosion of information on the Web and in easily accessible digital formats forces us to think “outside the box” when tackling data-intensive “Web-scale” problems. Researchers must think and analyze data at a massively parallel scale or face the prospect of being relegated to work on “toy” problems. “Cloud computing” could potentially provide the infrastructure that allows researchers to tackle “Web-scale” challenges at a reasonable cost. From an educational point of view, the ability to think about problems in terms of parallel processing algorithms will become a critical skill in tomorrow’s work force. “Cloud computing” is an emerging technology that iSchools cannot afford to ignore.

2. Goals

- To introduce the iSchool community to “cloud computing” and the MapReduce framework
- To provide the iSchool community an overview of research and education efforts currently underway
- To begin a discussion on the implications of “cloud computing” to research and education in iSchools

3. Proposed Format

I propose a 60 minute session structured in the following manner:

Overview of cloud computing [30 minutes]

(I will deliver this presentation)

- Description of the MapReduce framework.
- Discussion of the types of data-intensive information processing applications that MapReduce was designed for. I take “information processing” to broadly encompass information retrieval, natural language processing, text mining, social network analysis, etc.,
- A short demo of Hadoop, the open-source implementation of MapReduce.
- Overview of research projects at the University of Maryland that exploit this resource.
- Overview of the educational efforts associated with this initiative.

Panel commentary [15 minutes]

- Three panelists will be invited to share their views on what “cloud computing” could mean to research and education in iSchools.

Open discussion [15 minutes]

- The floor will be open to questions from the audience, and I will moderate a discussion.

3. Participants

The following panelists have been invited:

- John M. Unsworth, Dean and Professor, Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign. (participation confirmed)
- Another faculty from either an iSchool or from one of universities involved in the IBM/Google collaboration (to be arranged).
- Representative from IBM (to be arranged).

4. References

Selected media mentions of the IBM/Google initiative:

- New York Times: Google and I.B.M. Join in ‘Cloud Computing’ Research (October 8, 2007)
<http://www.nytimes.com/2007/10/08/technology/08cloud.html>
- University of Maryland Press Release (October 8, 2007)
<http://www.newsdesk.umd.edu/culture/release.cfm?ArticleID=1515>
- Washington Post: Maryland Joins Megacomputer ‘Cloud’ Project (October 9, 2007)
<http://www.washingtonpost.com/wp-dyn/content/article/2007/10/08/AR2007100801521.html>

Relevant technical articles:

Luiz Andre Barroso, Jeffrey Dean, and Urs Holzle. (2003) Web Search for a Planet: The Google Cluster Architecture. *IEEE Micro*, 23(2):22-28.

Jeffrey Dean and Sanjay Ghemawat. (2004) MapReduce: Simplified Data Processing on Large Clusters. *Proceedings of the 6th Symposium on Operating System Design and Implementation (OSDI 2006)*.

Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Michael Burrows, Tushar Chandra, Andrew Fikes, and Robert Gruber. (2006) Bigtable: A Distributed Storage System for Structured Data. *Proceedings of the 7th Symposium on Operating System Design and Implementation (OSDI 2006)*.

WILDCARD SUBMISSION 3:

Narratives of engagement: Working at the intersections of information, social, and domain science

AUTHOR(S):

Ribes, David

ORGANIZATION(S):

University of Michigan, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

‘Wildcard’ PROPOSAL – iCONFERENCE 2008

Narratives of engagement: Working at the intersections of information, social, and domain science

David Ribes
Workshop Organizer and Moderator
School of Information
University of Michigan
dribes@umich.edu

Christine L. Borgman
Department of
Information Studies
UC Los Angeles
finholt@umich.edu

Geoffrey C. Bowker
Center for Science,
Technology and Society
University of Santa Clara
bowker@usc.edu

Thomas A. Finholt
School of Information
University of Michigan
finholt@umich.edu

Tentatively Scheduled
Karen S. Baker
Scripps Institution of
Oceanography
UC San Diego
kbaker@ucsd.edu

OBJECTIVES

Engagement – sometimes called action research, contribution, or intervention – is the participation of the researcher in the object of study. It is a key contemporary feature in making iSchool research accessible and relevant to broader communities. However, it also presents novel challenges for the traditional academic endeavor. We have many ‘shorthands’ for these difficulties: misaligned reward structures, diverging goals, miscommunication across heterogeneous expertises, multiple membership, challenges of multidisciplinary collaboration, or unrealistic expectations on social science. 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. We have arranged this session to open and encourage discourse on the experiences, approaches and outcomes of engagement.

The core of the presentations will be centered on storytelling. Stories, or narratives, are ideal devices for capturing and conveying the complexities of real world field experiences. The four participants in this interactive panel will recount exemplary narratives of their engagement at the intersections of social research, information

studies and domain sciences. These stories 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 engaging social /information/domain science. This includes a range of activities stretching from policy recommendations, to participant observation or contributions to technology design; from distanced roles such as ‘shrink wrapped’ consulting, to long view historical studies, to becoming a daily members and ‘stakeholders’ in the success of projects.

STRUCTURE

This interactive session will be 1.5 hours in length.

The session will begin with participants’ brief presentations recounting a single exemplary experience of engagement and the research questions these activities have generated. We will then open the floor to discussion amongst presenters and with the audience.

The goal of this session are for the stories themselves to act as common starting points for a

collective discussion on the work (and pleasures) of engagement. Topics will emerge organically from discussion. This said, below are some of the topics we expect will come to structure the conversation:

- Venues for communicating approaches and findings back to our colleagues
- Traditions of 'objective' and 'subjective' research
- 'Going native'
- Methods or best-practices for collaboration
- The mythos of 'social engineering'
- Cautions in becoming 'a service science'
- Factors informing a good engagement
- Funding opportunities and dangers
- 'Outsourcing the social'
- Developing long-term partnerships

EQUIPMENT

Digital Projector.

White board or flip-chart

PARTICIPANTS

David Ribes is a postdoctoral research investigator. Trained as a sociologist at UC San Diego he currently holds a position at the University of Michigan, School of Information. Throughout his academic career David has worked with computer and domain scientists, studying and participating in the construction of large scale information infrastructure for the sciences (cyberinfrastructure). His dissertation research focused on the practical work of participants in the GEON project (cyberinfrastructure for the earth sciences) and today he continues this work through comparisons of projects for the geo- and environmental sciences.

Thomas A. Finholt is research professor and associate dean for research and innovation at the School of Information, University of Michigan. Tom's research focuses on the design, deployment, and use of cyberinfrastructure in science and engineering. He was a co-developer of the world's first operational collaboratory, the Upper Atmospheric Research Collaboratory (UARC), which was a finalist in the science category for the 1998 Smithsonian/Computerworld awards. His recent work has focused on the development of NEESgrid, the collaboratory component of the George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES). He has also conducted research on the impact of geographic dispersion and computer-mediated communication on trust and performance in virtual teams.

Geoffrey C. Bowker is Executive Director of the Center for Science, Technology and Society, Santa Clara University. His main research interests are in the field of classification and standardization: in particular asking how these play into the development of scientific cyberinfrastructure. His recent book *Memory Practices in the Sciences* looks at information infrastructures and storytelling in a science over the past two hundred years. It looks at geology in the 1830s, cybernetics in the 1950s and environmental sciences today - weaving together their information infrastructure and the stories that they tell about their objects. His work on information infrastructure involves looking at shifting classification systems in medicine, distributed collaborative work practices in environmental science, data sharing practices and biodiversity informatics. My central analytic question here is how scientists in the various sciences contributing to the subject of biodiversity communicate both with each other and with policymakers - and in particular how do the data structures and practices in use affect this communication.

Christine L. Borgman is Professor & Presidential Chair in Information Studies, UCLA.

Christine studies the creation, use, and management of scientific data and its implications for science policy. This research is associated with the Center for Embedded Networked Sensing and with the CENSEI Project. The goals are to apply knowledge of scientific data practices to the design of data collection and management tools, and to the design and policy of information services for research and Education. Christine's research on scientific data is one of many inputs to her second area of interest, which is analytical work on the changing nature of scholarship in an environment of ubiquitous computer networks and digital information. These threads are combined in her recent book, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*.

Karen S. Baker studies, designs, and practices within information environments. A member of UCSD Scripps Institution of Oceanography Integrative Oceanography Division, she is an information manager for the Palmer Station and California Coastal Ecosystem Long-Term Ecological Research sites. Karen is also a member of the Science Studies department at UCSD.

WILDCARD SUBMISSION 4:

Information Credibility: Let's Get Serious (and Have Some Fun Too!)

AUTHOR(S):

Eisenberg, Michael B

Lin, Peyina

Lin, Yuan

Lankes, R David:

ORGANIZATION(S):

University of Washington, United States of America

Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)



Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research

Miriam J. Metzger

University of California, Santa Barbara, Department of Communication, Ellison Hall, Santa Barbara, CA 93106-4020.

This article summarizes much of what is known from the communication and information literacy fields about the skills that Internet users need to assess the credibility of online information. The article reviews current recommendations for credibility assessment, empirical research on how users determine the credibility of Internet information, and describes several cognitive models of online information evaluation. Based on the literature review and critique of existing models of credibility assessment, recommendations for future online credibility education and practice are provided to assist users in locating reliable information online. The article concludes by offering ideas for research and theory development on this topic in an effort to advance knowledge in the area of credibility assessment of Internet-based information.

Introduction

In just the last two decades, the Internet has become integrated into our lives as an important, if not indispensable, tool for information and communication (Fallows, 2005). The plethora of information available online, coupled with heavy reliance on the Internet by information seekers raise issues of the credibility or quality of information found online. Credibility in this context refers to the *believability* of some information and/or its source (Hovland, Janis, & Kelley, 1953). A long history of research finds that credibility is a multifaceted concept with two primary dimensions: expertise and trustworthiness. Secondary factors affect credibility perceptions as well, including source attractiveness and dynamism, for example (O'Keefe, 2002). It also is understood that the credibility of a source or message is a receiver-based judgment which involves both objective judgments of information quality or accuracy as

well as subjective perceptions of the source's trustworthiness, expertise, and attractiveness (Freeman & Spyridakis, 2004).

Recent concerns about credibility stem from the fact that Internet and digitization technologies both lower the cost of and increase access to information production and dissemination. The result is that more information from more sources is available and more easily accessible now than ever before. In the past, substantial costs of information production and dissemination on a mass scale limited the number of sources to only those with enough authority and capital to justify and sell an information product. In the digital environment, however, nearly anyone can be an author, as authority is no longer a prerequisite for content provision on the Internet. This obviously raises issues of credibility, a problem that is exacerbated by the fact that many Web sites operate without much oversight or editorial review. Unlike most traditional (i.e., print) publishing, information posted on the Web may not be subject to filtering through professional gatekeepers, and it often lacks traditional authority indicators such as author identity or established reputation. Additionally, there are no universal standards for posting information online, and digital information may be easily altered, plagiarized, misrepresented, or created anonymously under false pretenses (Fritch & Cromwell, 2001, 2002; Johnson & Kaye, 2000; Metzger, Flanagan, Eyal, Lemus, & McCann, 2003; Rieh, 2002).

Burbules (1998) further suggested that because information is presented in a similar format online (i.e., Web sites), a kind of "leveling effect" is created that puts all information on the same level of accessibility, and thus all authors on the same level of credibility in the minds of Internet users. Studies of Web-based health information substantiate fears regarding the credibility of Internet-based information by concluding that the quality of online health information varies dramatically, with much of the information being inaccurate and incomplete (Eysenbach, Powell, Kuss, & Sa, 2002; Kunst, Groot, Latthe, Latthe, & Kahn, 2002; Morahan-Martin & Anderson, 2000; Rice & Katz, 2001; Seidman, 2006).

Received August 15, 2005; revised March 28, 2006, January 24, 2007; accepted January 25, 2007

© 2007 Wiley Periodicals, Inc. • Published online 21 September 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20672

The culmination of all this is that the Internet has made the need to critically evaluate information more important than ever before while also shifting the burden of credibility assessment and quality control from professional gatekeepers onto individual information seekers. Developing the skills to evaluate Web-based information, then, is crucial for Internet users; however, there is evidence that many people are unprepared for this responsibility and may have trouble determining how to assess the credibility of online information (Amsbary & Powell, 2003; Meola, 2004; Metzger, Flanagin, Eyal, et al., 2003; Scholz-Crane, 1998).

The aims of this article are to critically review current efforts to help Internet users acquire the skills needed to find credible information online and to evaluate existing theoretical models and user-based strategies of credibility assessment. The article proceeds by describing the common "checklist" approach to credibility assessment, which is then compared with other (e.g., cognitive and contextual) models of credibility, as well as empirical studies of users' evaluations of Internet information. Findings of this review are then leveraged to (a) suggest more effective strategies for educators and practitioners in helping information consumers discern credibility, (b) evaluate possible tools and social applications that could be developed to assist users in locating credible information online (e.g., seal programs, rating systems, vetted databases, digital signatures, and collaborative filtering), and (c) propose a research agenda for future work in this area, including a new "dual processing" model of credibility assessment.

Critical Evaluation Skills: Training Users to Evaluate Online Information

Shortly after the problem of establishing credibility in the online environment was recognized, efforts to educate and train users were under way. Many of these efforts were couched within the "digital literacy" movement and led by such groups as the American Library Association, the National Institute for Literacy (Kapoun, 1998; Rosen, 1998; Smith, 1997), and various healthcare agencies and consumer groups (Freeman & Spyridakis, 2004). A common aim of these groups was to assist Internet users in developing the skills needed to critically evaluate online information.

A key starting point for the digital literacy movement was the understanding that the skills needed to determine the quality or credibility of online information are largely the same as those for evaluating information found in other channels of communication (Alexander & Tate, 1999; Brandt, 1996; Fritch & Cromwell, 2001). Based on that, the literature identifies five criteria that users should employ in their assessments of the credibility of Internet-based information: accuracy, authority, objectivity, currency, and coverage or scope (see Alexander & Tate, 1999; Brandt, 1996;

Fritch & Cromwell, 2001; Kapoun, 1998; Meola, 2004; Scholz-Crane, 1998; Smith, 1997).¹

Accuracy refers to the degree to which a Web site is free from errors, whether the information can be verified offline, and the reliability of the information on the site. The *authority* of a Web site may be assessed by noting who authored the site and whether contact information is provided for that person or organization, what the author's credentials, qualifications, and affiliations are, and whether the Web site is recommended by a trusted source. *Objectivity* involves identifying the purpose of the site and whether the information provided is fact or opinion, which also includes understanding whether there might be commercial intent or a conflict of interest on the part of the source, as well as the nature of relationships between linked information sources (e.g., the meaning of "sponsored links" on a Google search output page). *Currency* refers to whether the information is up to date. *Coverage* refers to the comprehensiveness or depth of the information provided on the site. These recommendations require a range of activities on the part of users, from simple visual inspection of a Web site to more laborious information verification and triangulation efforts.

Checklist Approaches

Training programs based on these evaluative criteria typically develop checklists to guide users through the credibility evaluation process. Specifically, users are taught to ask and answer a list of questions designed to cover each criterion. For example, for currency, users are to check for a date stamp indicating when the information was originally uploaded or last changed (i.e., "Does the site provide information about when the information was last posted or updated?"). Under accuracy, they are to check if the author provides contact information (i.e., "Does the Web site list contact information such as a phone number or address?"). One problem with this kind of "checklist" approach is that it is rather time consuming and labor intensive for individual Internet users to perform for each site they visit (Meola, 2004).

Indeed, studies have found that users are seldom diligent in checking the accuracy of the information they obtain online (Flanagin & Metzger, 2000; Scholz-Crane, 1998). In a series of studies, Metzger and Flanagin examined the degree to which Internet consumers use each of the recommended criteria (i.e., accuracy, authority, objectivity, currency, and coverage) to gauge the credibility of the information they

¹Most digital literacy efforts focus on users' assessments of the credibility of Web pages, and so this discussion is accordingly limited to this form of online information; however, note that there are many other types of Internet-based information that have serious credibility implications, such as blogs, wikis, social networking sites, chat groups, and e-mail. Each of these forms of communication is somewhat unique and carries with it specific credibility concerns. That said, the general principles of critical evaluation should be applicable, albeit to different degrees, across all types of network-based information.

TABLE 1. Frequency of verification behaviors across each sample and overall.

Year data collected: Population sampled: <i>N</i> =	Dataset (<i>M</i>)					Overall <i>M</i>
	1999 students 718	1999 nonstudents 323	2000 students 485	2001 students 300	2001 nonstudents 274	
Check to see if the information is current.	3.18	3.21	2.68	3.16	3.66	3.18
Consider whether the views represented are facts or opinions.	2.88	2.95	3.02	3.10	3.61	3.11
Check to see that the information is complete and comprehensive.	3.04	2.91	2.90	3.09	3.26	3.04
Seek out other sources to validate the information.	2.57	2.82	2.85	2.79	3.38	2.88
Consider the author's goals/objectives for posting information.	2.65	2.69	2.39	3.03	3.15	2.78
Check to see who the author of the Web site is.	2.25	2.51	1.85	2.43	3.02	2.41
Look for an official "stamp of approval" or a recommendation from someone you know.	2.20	2.45	N/A	2.43	2.96	2.51
Check to see whether the contact information for that person or organization is provided on the site.	2.09	2.31	2.10	2.27	2.91	2.34
Verify the author's qualifications or credentials.	1.97	2.25	1.78	2.09	2.61	2.14

Note. 1 = never, 2 = rarely, 3 = occasionally, 4 = often, 5 = all the time.

Please see the following citations for detailed information about the sample and questionnaire items (Flanagin & Metzger, 2000, 2007; Metzger, Flanagin, & Zwarun, 2003).

found online. Survey data were collected across 3 years (1999–2001) from five different samples. (Three samples consisted of college students, and two samples were of general adult Internet users; $n = 274$ – 718). Respondents were asked to indicate how often they performed nine behaviors when visiting Web sites. Specifically, they were asked how often they check if the information is up-to-date (i.e., currency), to consider whether the views represented on a site are facts or opinions (i.e., objectivity), to consider the author's goals/objectives for posting the information (i.e., objectivity), to check that the information is complete and comprehensive (i.e., coverage), to seek other sources to validate the information on a site (i.e., accuracy), to check the author of the Web site (i.e., authority), to verify the author's qualifications or credentials (i.e., authority), to check whether the contact information for the author or organization is provided on the site (i.e., authority), and to look for an official "stamp of approval" or a recommendation from someone they know (i.e., authority).

Data analysis revealed striking similarity in both the frequency and nature of respondents' evaluation behaviors across all samples (see Table 1). Overall, respondents in every study reported performing each of the nine evaluation behaviors only "rarely" to "occasionally." Users evaluated Web sites' currency, comprehensiveness, and objectivity most often (although still only occasionally) whereas checking the author's identity, qualifications, and contact information were evaluated least often by respondents across the samples. Looking for other sources or recommendations and considering the goals of the author of the information fell in the middle. In general, Internet users scored highest on the actions that are easiest to perform and that only require their opinion (e.g., considering whether a site's information is current and complete), and lowest on the recommendations that are more time consuming and that require effort to

perform (e.g., verifying the qualifications or credentials of the author), even if the effort is fairly minimal (i.e., checking if contact information is provided). These data are interesting in many ways, but the most worrisome finding is that the strategy least practiced (i.e., verifying an author's qualifications) is perhaps the most important for establishing credibility.²

Another study using different research methodology similarly found that Internet users do not vigorously apply the five criteria in their judgments of information quality (Scholz-Crane, 1998). In this study, 21 college students evaluated two Web sites. Participants were asked to write in an essay format "how they would evaluate the quality of each site including specific criteria used to assess this quality" (p. 55). Content analyses revealed that of the five criteria, most students used only two in their evaluations of information quality: scope (e.g., the site provides detailed information) and accuracy (e.g., the site contained statistics, cited sources, and was clearly written and well organized). Only 6 students assessed objectivity by considering whether there was evidence of bias on the sites, and fewer considered authority (by looking at the author identity or site sponsor) or contact information. In many cases, students used a single criterion in their final decision about the quality of the

²Across all datasets, evaluation behavior was positively related to experience online. At first glance, this bodes well for the future: As people gain experience, they will be more likely to evaluate online information; however, there is some evidence that people with greater experience exaggerate their critical evaluation behavior on self-report measures (see Flanagin & Metzger, 2007). In addition, education seems to be a factor, as people with more education report evaluating online information more often than do those with less education. This speaks to the need for digital literacy efforts to target at-risk groups, including younger and less educated populations.

Web sites. A clear conclusion from the various Metzger and Flanagin studies and from the Scholz-Crane (1998) data is that few users are rigorously evaluating the quality of the information they obtain via the Internet (also see Eysenbach & Kohler, 2002; Wilder, 2005).

Additional Models of Web Information Evaluation

The checklist approach to credibility assessment is the most popular among educators, although other models have been proposed. For example, Fritch and Cromwell (2001, 2002) presented a model for ascribing cognitive authority to Internet information. Cognitive authority is a term they used to incorporate both the notions of credibility and quality; it is what people evaluate "to know what credence and impact to impart to any specific body of information" (2002, p. 243). Fritch and Cromwell (2001) proposed an iterative model whereby assessments of authority and credibility are made by the information seeker at the levels of author, document, institution, and affiliations, which then are integrated into a global judgment of online information credibility. In this model, verifying the author and/or institutional identity of a Web site through reputation or stated qualifications, considering the factual accuracy of the Web document and its presentation and format, and examining both overt and covert affiliations of the Web site are recommended strategies that, when combined, contribute to an overall impression of its cognitive authority.³ This model is similar to the previously described "checklist" model in that it provides a "to do" list for Internet users, but differs from that model in that it places more emphasis on the technological tools available to users for making these assessments. For example, to help establish author identity and site affiliations, Fritch and Cromwell (2003) recommend using tools such as Whois, Traceroute, and NSlookup/Dig. These technologies can assist users in finding the identity of the owner of a particular Web site and can reveal affiliations between organizations or Web sites that are not immediately apparent.

Walther and Burkell (2002) also proposed an iterative model for how users judge the credibility of online information based on a review of literature in psychology and communication. They posited that credibility assessment takes place in three stages. First, users form an impression of the overall site credibility by examining its surface characteristics including, for example, its appearance and presentation (e.g., colors, graphics, typographical errors), usability and

interface design (e.g., navigability), and organization of information. Next, the information or site content is evaluated for its credibility by looking at characteristics of the source (e.g., expertise, trustworthiness, credentials) and message (e.g., currency, accuracy, relevance to the user). The third stage of their model involves factoring in the user's cognitive state at the time of evaluation. In other words, assessments of the message presentation and content will differ depending on the users' need for information, need for cognition, and prior knowledge of the topic, and these will interact with other situational and individual-level factors (e.g., time available for information retrieval and processing).

This model resembles Fritch and Cromwell's (2001) model for ascribing cognitive authority in that it suggests there are different levels of credibility assessment that are combined to form a final judgment (also see Rieh, 2002). Another similarity to the checklist model is that it provides evaluative criteria that users are supposed to consider in their credibility assessments. The model differs from the others, however, in that it incorporates aspects of the information receiver as being important to the evaluation process, highlighting the fact that credibility judgments are situational and dependent on individual-level factors. Indeed, information scientists have long understood that both a user's informational needs and context are fundamental to information evaluation and use.

Factors Important to Users of Web-Based Information

One problem with the foregoing models of credibility evaluation is that none has been tested using actual Internet users as they seek information online. Another problem, related to the first, is that these models have not been directly connected to studies examining what factors people *really do* use to determine credibility online. Researchers have suggested myriad factors that may play into credibility assessments (see Table 2), but only a few studies have examined what criteria people actually employ. The studies that do so are reviewed next (see also Morahan-Martin, 2004).

In a study involving 21 participants, Eysenbach and Kohler (2002) examined how consumers appraise the quality of online health information. The research was conducted in three stages, beginning with focus groups to identify the criteria participants say they use in deciding whether Internet health information is credible. The second stage involved observing a subsample of the focus-group participants as they sought health information online that they "felt confident about." The third stage consisted of the researchers interviewing them about their searches to understand participants' decision-making processes and criteria for selecting particular Web sites. Interestingly, what focus-group participants said they looked for in assessing credibility was not what the researchers found they actually looked at during the observational portion of the study. An analysis of the focus-group discussions revealed that participants' main criteria for assigning credibility rested on whether the source was an official authority, whether the page cited scientific references, whether the site was professionally designed and easy

³Seidman (2006) offered a model of third-party health Web site evaluation that involves experts vetting and rating sites along structural and performance criteria. Structural criteria include whether a health site: (a) explains the process by which content was generated and selected for inclusion; (b) provides author identity, affiliations, credentials, and contact information; (c) uses peer review and supports claims with evidence; and (d) regularly dates and updates content. Performance criteria include measures of the comprehensiveness and accuracy of the information provided on the site against known evidence-based practice guidelines and expert medical opinion.

TABLE 2. Suggested factors that influence credibility assessments of online information.

Presence of date stamp showing information is current
Source citations
Citations to scientific data or references
Author identification
Author qualifications and credentials
Presence of contact information
Absence of advertising
Presence of privacy and security policies
Certifications or seals from trusted third parties
Professional, attractive, and consistent page design, including graphics, logos, color schemes, etc.
Easy navigation, well-organized site
Sponsorship by of external links to reputable organizations
Notification/presence of editorial review process or board
Absence of typographical errors and broken links
Professional-quality and clear writing
Download speed
Message relevance, tailoring
Interactive features (e.g., search capabilities, confirmation messages, quick customer-service responses)
Past experience with source/organization (reputation)
Domain name and URL (suffix)
Ability to verify claims elsewhere (e.g., external links)
Comprehensiveness of information provided
Ranking in search engine output
Paid access to information
Plausibility of arguments

Source: Alexander & Tate, 1999; Eysenbach & Kohler, 2002; Fogg et al., 2003; Freeman & Sprydkis, 2004; Metzger et al., 2003; Rieh & Belkin, 1998; Rieh, 2002; Walther & Burkell, 2002.

Note. Of course, not all factors will apply to every Web site. Site type (e.g., a commercial vs. an individual's Web site), and receiver factors (e.g., purpose and motivation for information search, Internet experience and prior knowledge) also will make a difference in which factors are applicable and which factors are used during the evaluation (Flanagin & Metzger, 2000; Fogg et al., 2003). In addition, some factors are more or less helpful at evaluating credibility at different "levels" (e.g., at the level of the Web site as a whole or at the level of messages residing on Web sites). To illustrate, factors having to do with site design and navigation will be more relevant to judgments of site credibility whereas factors such as argument plausibility and author credentials will likely be more important in evaluations of the credibility of messages on Web sites.

to use, and whether it was written in language that they could understand. Despite placing emphasis on the identity of the source and quality of the data presented on a Web site, the observational portion of the study revealed that while most participants used sites they were unfamiliar with, none looked at information about the source or how data on the sites were compiled. Perhaps most distressing, the post-search interviews revealed that "few participants took notice and later remembered from which websites they retrieved information or who stood behind the sites" (p. 576).

Eysenbach and Kohler's (2002) findings overlap somewhat with those of a study of online health-information seekers conducted by the Pew Internet and American Life project (S. Fox & Rainie, 2002). Using surveys and focus groups, this study found that overt commercialism, trusted third-party endorsements, site-design elements, as well as users' ability to determine the source and currency of the information posted on health-related Web sites were the most important criteria for assessing the credibility of online health information.

Rieh (2002) examined people evaluating the quality of information as they performed a variety of search tasks.

Using think-aloud and interviewing procedures, she followed 15 academics as they searched for information online regarding four topics: academic research, health, travel, and product (ecommerce) information. Each participant was instructed to find information on Web sites that they considered to be of high quality. Rieh found that in making evaluations, participants' judgments were based on two basic criteria: the characteristics of the information "objects" that they found online and the characteristics of the source of online information. The characteristics of the information objects that participants were concerned about included the type of information object (e.g., journal article vs. chat forum), its content (i.e., how detailed, comprehensive, and technical the information provided is), its presentation (i.e., graphics), and its structure (i.e., how well organized the information is). Source characteristics that participants looked to were reputation and type of source (e.g., commercial vs. noncommercial source, .com vs. .edu, etc.).

A problem with studies such as those just reviewed is that they tend to rely on small, atypical user populations performing somewhat specialized search tasks. As a result, it may be difficult to generalize findings to other users. By contrast,

Fogg et al. (2003) conducted a large-scale study of a variety of user populations. In an online experiment, 2,648 people evaluated the credibility of several real Web sites that ranged in their informational content (e.g., news sites, ecommerce sites, nonprofit sites, health sites, travel sites, etc.). Fogg et al. analyzed user comments "to find out what features of a Web site get noticed when people evaluate credibility" (p. 1). Not surprisingly, results varied by the type of Web site evaluated (e.g., ecommerce sites vs. news sites, etc.), and most respondents mentioned many features. Summarizing the findings, people seemed to consider four broad categories of elements. The first was *site presentation*. This included visual design elements such as the graphics and readability of the site as well as the site's navigability and functionality (e.g., broken links, presence of a search feature). Interestingly, site design/presentational elements were the most frequently used criterion, with 46% of the comments including this feature. A second consideration related to the *information* on the page, including its organization, breadth/depth, accuracy, bias, usefulness, and the tone and clarity of writing. Third, the site operator's or source's *motives* were a factor. This includes selling intent, presence of advertising, clear identification of the site sponsor, and treatment of customers. Finally, the source's *reputation* played into users' credibility judgments, including name recognition, their past experience with the site, presence of seals of approval, or affiliations with reputable organizations.

To summarize, the results of the Fogg et al. (2003) study are similar to those of Rieh (2002) and Eysenbach and Kohler (2002) in that features of both the information itself (e.g., its organization, level of detail, etc.) and the source (e.g., commercial intent) matter in Internet users' credibility assessments. Results also reflect Scholz-Crane's (1998) finding described earlier that not all five of the critical evaluation skills criteria are used. Indeed, the data from Fogg et al.'s study show that people only used four of the five criteria: authority (e.g., through source reputation), accuracy of information, comprehensiveness of information, and objectivity (e.g., in evaluations of information bias and source motive). Together, this collection of studies indicates that people do not seem to take the currency of the information they find online into account when making credibility judgments. This is somewhat surprising, given that many Web pages indicate when the information was last updated at the bottom of the page.⁴ In addition and perhaps most interesting, the review of research reveals that people rely most heavily on a criterion that is not among the five critical evaluation skills recommended to judge credibility. That is, design/presentational elements appear to be the primary factor in users' credibility and information-quality assessments. The

implications of this finding are ominous in that they suggest Internet users may be easily duped by slick Web design. Taken together, these studies speak to a need to think beyond the checklist model to get people to critically evaluate the information they find online.

A Contextual Approach to Credibility Assessment

A very different approach to the checklist model and other models offering evaluative criteria lists is Meola's (2004) contextual model of Web site evaluation. Meola critiqued the idea of using checklists to evaluate online information because he believes they are unwieldy for users to perform (e.g., one checklist requires Internet users to answer over 112 questions per Web site viewed), and thus are an unrealistic approach to teaching critical evaluation. The aforementioned research (e.g., Flanagan & Metzger, 2000; Scholz-Crane, 1998) showing people's reluctance to exert much effort to verify online information supports his view. He also challenged the notion that all online information needs to be verified to the same degree by noting that more Web-based information these days is professionally vetted. In fact, he made a useful and important distinction between the "free Web" and the "fee-based Web" when it comes to credibility assessment. Although much of what is available on the free Web may not be subjected to peer or editorial review, the situation is very different for the fee-based Web, which includes professionally vetted subscription databases, prestigious journals, or national newspapers offering full-text access to their contents.

In contrast to the checklist model with its emphasis on evaluating the internal characteristics of Web sites (e.g., identifying the author, checking when the information was last updated, etc.), the contextual model focuses on information external to a particular site. By using external information to establish credibility, Meola (2004) contended that online information "is located within its wider social context, facilitating reasoned judgments of information quality" (p. 338). Meola's approach recommends three techniques to determine the quality of online information. The first technique is (a) *promoting peer- and editorially reviewed resources* that are available online. Information intermediaries (e.g., teachers, librarians, experts, etc.) should inform Internet users of high-quality, vetted resources that are searchable through universities, schools, public libraries, by subscribing to national newspapers online, or through other (largely fee-based) portals. This is similar to when doctors recommend Web sites to patients as a sort of "information prescription" (Rice & Katz, 2006). Meola noted that the time currently taken by teaching Internet users a checklist of questions is better spent teaching people what high-quality resources are available and how to use them.

The second technique is *comparing* information found on a Web site to other Web sites and/or to offline sources such as, for example, newspapers or magazines, peer-reviewed journal articles, or books. This is a more practical strategy for the "free Web" and, as Meola wrote: "Comparing Web

⁴It also is surprising because it contradicts Metzger and Flanagan's consistent pattern of finding that checking if the information is current is the strategy used most often across their five samples. The discrepancy in results could be due to the methods used by different researchers (e.g., self-report vs. direct observation), which underscores this article's ultimate conclusion that the study of online credibility must be conducted using a variety of research techniques.

sites to each other and to reviewed sources provides an understanding of the depth of information available, reveals the controversial areas of a topic that need special attention, and gives . . . a feel for the different kinds of information available across formats" (p. 341). The final technique is *corroboration*. Similar to comparison, corroboration involves seeking more than one source to verify information on a given topic. Internet users may assess the accuracy and reliability of information through corroboration as a convergence of facts and/or opinions from a variety of sources is (or is not) found. In the end, Meola argued that the contextual approach to online information evaluation is more practical than are the checklist approaches, and thus will be more likely to be used by online information seekers.

Implications and Recommendations

The foregoing review leads to a number of conclusions and recommendations about Internet credibility and the user. In particular, it suggests new strategies for practice and/or for curriculum development, and new technological and social approaches to helping users assess the credibility of the information they find on the Internet. It also suggests several directions for future research and opportunities for theory development on user assessments of the credibility of online information.

Educational and Practical Efforts

Although the checklist approach to critical evaluation is the most popular among educators, it is perhaps not the most practical or effective for the reasons described earlier. The research has clearly indicated that the best strategy for educating Internet users will be one that is easy to use and/or is transparent to them. Indeed, a closer connection between research findings on credibility assessment using actual online information seekers and curriculum development is long overdue. That means if educators continue to teach checklists for Web evaluation, they need to keep them short. In addition, the models of credibility assessment reviewed earlier suggest that the best strategy for practice is perhaps to teach a hybrid approach to credibility assessment that is based on individual Internet users' motivation and purpose for seeking information online. In other words, a variety of approaches to credibility assessment could be taught to fit the specific search task or situation at hand. Users could be taught to use the checklist and contextual models of credibility assessment in situations where they feel motivated to obtain high-quality, credible information such as when the negative consequences of obtaining incomplete or inaccurate information are high. For less motivated users or search situations, users could be taught some simple heuristics to consider, such as checking for source or sponsorship information. Of course, learning when to apply which approach would then need to become part of the educational effort. The advantage of this kind of "sliding scale" approach to teaching critical evaluation skills is that it is more focused and less

effortful for users to perform in the majority of their searches, and thus is more realistic for educators to expect of users.

Another conclusion from the review is that digital literacy efforts might be more effective if they are targeted at certain user populations who are particularly vulnerable for receiving low-quality information online or those whose needs for credible information are very high. Young students using Web-based information to learn or medical patients seeking information about their treatment options are examples. Information about credible sites and how to access high-quality databases residing on either the "free" or "fee" Web would be particularly valuable to these user groups and could be tailored by topic or area (e.g., medical, news, academic, etc.).

Working through schools and libraries is one very useful way to disseminate this kind of information to target user populations, but what about other ways of reaching users? Digital literacy efforts need to be extended outside the classroom and library reference desk. While literally dozens of online guides to Web site evaluation currently exist, a comprehensive Web site developed to serve as a resource for users to understand Internet credibility and how to assess it would be more useful.⁵ The site could have many levels of information and instruction regarding online credibility assessment, and users could go as deeply into the site as they wished depending on their motivation and interest in online credibility. In addition to general information about online credibility, the site could include specialized tips and instructions for seekers of particular kinds of information, such as ways to determine the credibility of medical and health information, commercial information, political information, and so on. Note that this type of effort would only succeed to the degree that it was well publicized and linked to several popular "point-of-entry" Web sites, which includes major portal sites that people typically use to launch their online information searches (e.g., popular search engines, medical information clearinghouse sites, government sites, etc.).

Social and Technological Means of Online Credibility Assessment

One interpretation of the research reviewed in this article is that credibility assessments should *not* be up to users because they are unlikely to exert the effort it takes to verify the credibility of Internet-based information.

⁵Examples of online Web site evaluation pages include *The Good, the Bad, and the Ugly, or Why It's a Good Idea to Evaluate Web Sources* produced by New Mexico State University (<http://lib.nmsu.edu/instruction/eval.html>), Kathy Schrock's *Guide for Educators* (<http://school.discovery.com/schrockguide/eval.html>), the Illinois Mathematics and Science Academy's 21st Century Information Fluency Project (<http://21cif.imsa.edu/>), among many, many others. The Credibility Commons, a joint project currently under development by the University of Washington's Information School and the Information Institute at Syracuse University, is an example of a credibility "clearinghouse" (<http://credibilitycommons.org/>) which will eventually offer a range of credibility tools and information to the public.

This suggests that energy may be better spent developing tools or systems that are designed to assess credibility for users, and then training users how to use those tools and systems. Several possibilities exist, some of which are described next (for a description of these and other possibilities, along with a discussion of the problems and challenges associated with each, see Fritch, 2003).

Credibility seal programs modeled after the TRUSTe or BBB online seal programs could be set up to assist Internet users in locating Web sites whose content has been approved by a trusted outside authority. A credibility seal program would capitalize on the fact that many people assess credibility based on somewhat-quick visual inspections of Web sites (Fogg et al., 2003). Such a "trustmarking" program would require the development of a code of conduct that sites carrying the seal would be required to adhere to (e.g., seal-bearing sites must provide authorship and update information). The HONcode program developed by the NGO Health on the Net Foundation is an example of a credibility seal program within the medical/health field (Fritch, 2003). According to its mission statement, the HONcode aims to "hold Web site developers to basic ethical standards in the presentation of information" and "help make sure readers always know the source and the purpose of the data they are reading" (see <http://www.hon.ch/HONcode/>). TRUSTe and HONcode provide viable models for credibility seal programs that could be extended beyond ecommerce and online health information.

Credibility rating systems could be instituted to help establish the credibility of Web-based information. Such systems could be controlled by an institution (e.g., the American Library Association) or could be based on a peer-rating system (a discussion of collaborative filtering via social networks appears later in this article). In any case, rating systems could be created that would allow Web sites to be assessed systematically along several quality standards (e.g., authority, currency, objectivity, disclosure, etc.). Peer-based credibility rating systems could be developed using interactive software that allows Web users to answer a series of questions about sites they visit and then assigns a rating based on aggregating users' responses.⁶ The ratings would be stored in a database and would be accessible to Internet users at a central location, or the ratings could be required by

government regulators to be displayed on all sites or integrated into all Web browsers. An application could be developed and downloaded that would post a small icon with a site rating, using the central rating database, which would not require the user to go get a rating for each site visited. Alternatively, a ratings system could be developed by having a panel of experts rate various sites. Due to the incredible volume of sites that would need to be evaluated, this may be easier to do within specific types of online content domains, such as health information (e.g., Seidman, 2006). The ratings systems described here are specific to assessing credibility. Social networking approaches more generally (e.g., reviews of books or magazines on Amazon.com, eBay feedback ratings, or MySpace friend links) can provide credibility cues as well, although they are not aimed to provide credibility ratings, per se. Social networking approaches to credibility assessment will be discussed later as collaborative filtering and peer review.

Directories, databases, or search engines that carry or display online content that has been preapproved or filtered for credibility could be developed. Many such proprietary databases already exist in the form of Lexis-Nexis, JSTOR, ERIC, and the like. For the so-called "free Web," it is possible that nonprofit groups or education associations could develop and implement searchable databases or directories of information from high-quality sources, based on some agreed-upon and publicized criteria. Examples are the Librarian's Index to the Internet (www.lii.org), which provides a list of librarian-vetted Web sites on various topics, and government-run health information portals such as Medline Plus (see <http://medlineplus.gov/> for an analysis of commercial vs. government online health databases, also see Rice, Peterson, & Christine, 2001). Google itself presents an interesting case in this regard because one of the reasons for its immense popularity and influence is precisely because it has a sort of built-in credibility metric: the number of inward links from other sites with a large number of inward links. Regardless of who develops or sponsors these databases, it is critical that they are easily searchable. High-quality databases such as those found on the "fee Web" and in libraries should be as easy to search as popular search engines such as Google.⁷ Reducing the effort involved in using high-quality, professionally vetted databases will increase their usage, which could in turn drive down user access and subscription costs.

Combining the last two ideas (i.e., credibility ratings and vetted databases), some propose using *Platform for Internet Content Selection (PICS) labels* to establish the credibility of Internet information (Eysenbach & Diepgen, 1998; Fritch, 2003). The existing PICS developed by the

⁶Fritch (2003) described two applications that could serve as prototypes for such a system. One application is the Information Quality (IQ) tool (<http://hitweb.mit.edu/mitweb.org/iq/>) that was developed to support the Health Summit Working Group's criteria for assessing the quality of health information on the Internet. The IQ tool is essentially a pop-up window that prompts the user to answer credibility-related questions about a specific Web site that he or she is currently viewing. Based on answers to these questions, the IQ tool calculates a "quality" score for the Web site. The other application works similarly, although it is not interactive. It is the DISCERN questionnaire (<http://www.discern.org.uk/>), which prompts users to rate health-related Web sites by answering focused questions pertaining to the quality of information found on the sites. In either application, ratings data could be collected across individuals over time to produce aggregate ratings.

⁷Interestingly, Google itself is beta testing an incarnation of the idea of providing searchable, vetted databases. Google Scholar (<http://scholar.google.com>) searches a specialized database that indexes articles from academic sources including academic publishers, professional societies and associations, universities, preprint repositories, and scholarly articles posted on the Internet.

World Wide Web consortium could be configured to filter and select information that meets certain credibility standards. PICS enables Web site operators to label their content, and allows users to control the content they receive based on those labels. If Web site operators (or others) included information in the labels about the nature and quality of information provided on the site, Internet users could use the labels to select information that meets their minimum criteria for credibility while also filtering out information falling below that minimum. More plausibly, third parties could rate or preapprove content along some credibility metric and use the PICS platform to create electronic labels for Web sites that reflect their evaluations. The labels then could be used by end users to filter information. The MedPICS Certification and Rating of Trustworthy Health Information on the Net (medCERTAIN) project is an example of such a system (Eysenbach, 2000; Eysenbach & Diepgen, 1998). The MedPICS system would work by labeling health-related information according to core standards of quality developed by medical societies, associations, and relevant others, and then entering that information into one or more medical-label databases that would be searchable by consumers.

More recently, Google has begun developing an algorithm that will rank and display search results on its all-news site, Google News (<http://news.google.com/>), according to the credibility of the news source. Currently, results are ranked according to criteria such as relevance to search term and date to determine which stories appear first, but the new algorithm would use such credibility-related factors as the news source's number of years in business, staff size, number of bureaus the news source operates, circulation information, average story length, volume of network traffic to its site, and other criteria to rank stories (Baertlein, 2005; B. Fox, 2005).

Digital signatures could be used to assure the authorship of online information, which is a key aspect of credibility. Fritch and Cromwell (2002) explain that digital signatures are "a cryptographic method that allows a sender to append a small data block, called a 'signature,' that allows the recipient to prove, to some quantifiably high certainty, two things: first, that message contents have not been altered in transit, and second, that the message sender is really who he/she claims to be" (p. 252). Authenticating the identity of the author is an important first step in assessing the credibility of online information, and is particularly important for Internet transactions involving sensitive information such as one's financial or personal data. Ultimately, however, digital signatures still require the recipient of some information to evaluate the quality of the information provided by the author (Fritch & Cromwell, 2002). This approach also is problematic for Web sites that are coauthored by more than one source, corporate authored, or provide no author information, as is increasingly the case. Indeed, any approach to credibility assessment that relies on author identity may be impractical as source information becomes more complex and difficult to pinpoint online.

Of course, many and even most of these solutions may be infeasible due to their high cost, low profitability, reliance on voluntary compliance, and tremendous implementation effort. In the end, Internet users may be left with no choice but to rely on themselves. If that is the case, then extending existing models of *collaborative filtering* and *peer review* of Web-based information, such as those found on *epinions.com*, *BizRate*, *eBay*, *Amazon*, or even in social networking sites such as *MySpace* or *Facebook*, might be a more practical way for users to discern whether some information they find online is credible. Peer review allows Internet users to provide feedback about the products, information, or source of some information offered on Web sites. It allows users to pool their intellectual and experiential resources when evaluating the trustworthiness, reliability, or quality of a Web site or information residing on a site, making credibility assessment a collective rather than an individual task. Hence, the name "collaborative" filtering (e.g., Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994).

User review and feedback systems, which also are known as recommender or reputation systems, were initially developed in the realm of ecommerce but now exist within several other Web content domains (e.g., online recipe sites such as *epicurious.com* use these systems). Indeed, at least 26% of adult Web users in the United States reported having rated a product, service, or person online (Rainie & Hitlin, 2004). Of course, peer-review systems are susceptible to biased, uninformed, or inaccurate reviews, as was demonstrated by revelations that *Amazon.com* book authors were promoting their own work by submitting anonymous reviews. The site *Slashdot.org* provides an interesting model of user review that tries to overcome the problem of bias in peer review. *Slashdot* is a large Internet forum in which users summarize information they find on other Web sites, link to those sites, and provide a place for readers to comment on the information. Contributors and their contributions are then rated by *Slashdot* users in a very elaborate and sophisticated system that helps to ensure fairness and impartiality. *Slashdot's* system of peer review covers a huge range of Web content and offers a provocative model that could be used to develop systems of peer review to assist Internet users in assessing the credibility of a wide variety of online information.

A Research Agenda

The review of literature presented earlier also clearly indicates that there is much more work to be done to formulate effective strategies for building Internet users' skills to critically evaluate online information. A research agenda for the issue of Internet credibility must include studies of information evaluation using a greater variety of research methods, on a greater variety of Internet users, performing a greater variety of search tasks than has been done to date. The body of findings from this research agenda should be used to develop and shape practical guidelines for Internet users, creating a strong linkage between research and practice. As part of this, the research agenda also must include evaluative

studies of the usability and effectiveness of any practical guidelines developed to help Internet users discern credible information online.

An area of immediate concern for scholars studying Internet credibility is the need for more research on what users actually do to assess credibility. This is true not simply because there are very few studies to date that have examined large populations of Internet users, but also because there is evidence that those studies may already be out of date. Specifically, existing research has found that more experienced Internet users may act differently than do less experienced users when it comes to credibility assessment (Freeman & Spyridakis, 2004; Johnson & Kaye, 2000, 2002; Flanagin & Metzger, 2007). This suggests that as users gain experience with this relatively young medium, perceptions of the credibility of information found online, as well as the ways in which users make their credibility assessments, may change. For example, Flanagin and Metzger (2007) found that although more experienced Web users *said* they evaluated information online more than did less experienced users in the study, they *actually* evaluated information less.

This discrepancy in self-reported versus actual evaluation behavior raises an important point with regard to the study of credibility online: The method of study may, by itself, influence the results obtained (also see Eysenbach & Kohler, 2002). Credibility research has the potential for response bias problems, as people know they "should" critically analyze the information they obtain online, yet rarely have the time or energy to do so. Given this situation, researchers must look to other methods besides survey questionnaires to understand actual online behavior. Research methods for future online credibility research should be as anthropological, naturalistic, and unobtrusive as possible.

Perhaps the most pressing item on the agenda for online credibility research is the need to study the role of user *motivation* in the credibility-assessment process. There is much evidence that user motivation, stemming in part from the information search task, moderates the degree to which users will critically evaluate online information. Surprisingly, however, user motivation is missing from most existing models of credibility assessment. The basis for this recommendation is the observation that not all Web sites need vigorous examination (Fritch & Cromwell, 2001, 2002; Meola, 2004), and research finding that different evaluative processes are used for different search tasks (Fogg et al., 2003; Rieh, 2002). As Fritch and Cromwell (2002) wrote:

Sometimes the significance of the information under scrutiny does not warrant extensive research and analysis, while in other instances extensive research and analysis may be crucial; it ultimately depends upon the purposes and intent of the user(s) of the information." (p. 251)

This further suggests that credibility assessment should not be conceptualized as a single evaluative process. Instead, different users may use different processes at different times to evaluate the credibility of online information.

Dual processing models of information processing and assessment may be a helpful guide in formulating a model of credibility assessment that takes motivation into account. Dual processing models are proposed to operate in many areas of social and cognitive psychology, and have been successful in explaining a wide array of social judgments including, for example, persuasion, impression formation, person perception, and decision making (Chaiken & Trope, 1999). Examples of some well-known dual process theories involving information assessment include the Elaboration Likelihood Model, the Heuristic-Systematic Model, and controlled versus automatic processing models (Chaiken, 1980; Petty & Cacioppo, 1981; Shiffrin & Schneider, 1977), although there are others (for a review, see Chaiken & Trope, 1999).⁸

What the various dual processing models have in common is an emphasis on the role that motivation and cognitive ability play in guiding information assessment and decision making. In brief, these models theorize that people will process and/or scrutinize messages in more or less depth depending upon the receiver's motivation and ability to do so. In other words, when people are motivated due to personal or situational factors such as having a high need for accurate information or having a personal stake in understanding some issue, they are likely to pay more attention to a message, think about the arguments presented, and exert more cognitive resources to process and evaluate the information and its source. By contrast, when motivation and/or ability are low, these models predict that information will be processed or evaluated based on more superficial and less thoughtful criteria. In these situations, decisions will be made on more heuristic judgments of the message or its source (e.g., attractiveness) rather than on message quality. Ability to process a message hinges on many factors (e.g., sufficient time, prior knowledge of the topic, message comprehensibility, etc.).

A Dual Processing Model of Credibility Assessment

Dual processing models provide a good basis for developing a new model of Web credibility assessment that prioritizes user motivation and ability and accounts for the situational nature of credibility assessment (i.e., that it will not be important for all Internet users at all times). Like most dual processing models, the model proposed in Figure 1 assumes that motivation and ability are keys to whether and to what degree users will critically evaluate Web information. In this context, motivation stems from the consequentiality of receiving low-quality, unreliable, or inaccurate information online. Ability may be linked to users' knowledge about how to evaluate online information, which could involve teaching users the critical evaluation skills employed by the checklist or contextual models of credibility, for example.

⁸Indeed, several researchers have discussed the Elaboration Likelihood Model with regard to online credibility processes (see Fogg et al., 2003; Freeman & Spyridakis, 2004; Walther & Burkell, 2002).

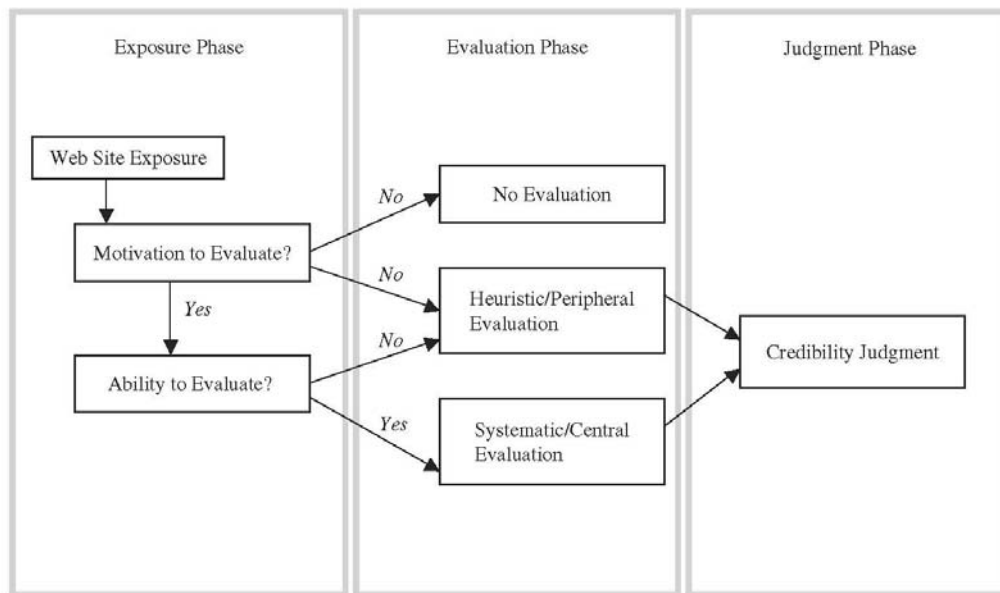


FIG. 1. Elements of a dual processing model of Web site credibility assessment.

Simply put, the degree to which online messages will be scrutinized for their credibility depends on individual users' ability to evaluate the message and their initial motivation or purpose for seeking the information, which involve their awareness of and salience of the consequentiality of the information.

In dual processing theory terms, Chen and Chaiken (1999) stated that perceivers' "accuracy goals" may vary depending on the situation such that people are more or less motivated to reach accurate judgments across different contexts. Applied to the online environment, while Internet users are active in searching for information online, their accuracy goals may vary from search to search. Furthermore, Internet information seeking may be more or less "casual" or "purposeful" depending on the situation. While we can assume that some Web browsing is highly motivated by a person's need to find accurate information, a good deal of users' online information-seeking behavior is not so purposeful. For example, a person may go online for casual-entertainment purposes, begin a search on one topic but be lead to other information via hyperlinks, or be confronted with information he or she did not intend to find when opening a new page. Despite being less purposeful, there is no reason to assume that users in these scenarios would not be concerned with the credibility of the information they find online. By the same token, one might assume that they may be *less* concerned about credibility, and thus are less willing to exert maximum cognitive resources toward credibility assessment.

A dual processing model of credibility is not only able to predict *when* users will or will not make the effort to critically evaluate online information but also is useful in understanding *how* credibility assessments are made (i.e., what processes are involved in those assessments). A dual processing model of credibility assessment would propose that users look to different aspects of Web-based information to assess its credibility depending on their motivation for seeking information. Users who are less motivated to find high-quality, credible information (e.g., a person biding some time online by looking up information on a favorite sports team or music band) may not assess credibility at all or do so by simply considering such simple characteristics or heuristics as a Web page's design and graphics. On the other hand, highly motivated users will likely take a more rigorous, systematic approach to credibility assessment. Examples might include a medical patient recently diagnosed with a serious disease, a student preparing a research paper, or a business owner analyzing the market. In these cases, users are more likely to go beyond the surface characteristics of a Web site to examine the information content, its author or source, and other factors in determining its credibility (for similar arguments, see Fogg et al., 2003; Walther & Burkell, 2002).

One way to think about this in light of the previous discussion is that the proposed dual processing model can predict when users are likely to use a relatively more effortful credibility assessment procedure (e.g., the checklist approach) versus when they are likely to use an approach that

is potentially less time-consuming (e.g., the contextual approach) or perhaps even a very simple heuristic approach (e.g., simply looking at the site's graphic design) to determine the credibility of some online information. Another way that this new model of credibility assessment is useful is that it can help to understand when people might evaluate different types of credibility. Several scholars have suggested that users assess various types or "levels" of credibility in their judgments of information quality and trustworthiness (Fritch & Cromwell, 2001; Metzger, Flanagin, Eyal, et al., 2003).⁹ Fogg (2003), for example, discussed four types of credibility that Internet users assess: presumed, surface, reputed, and earned. The dual processing model of credibility assessment could help to predict when users are likely to simply look at the surface credibility of a site (e.g., simple design or layout) and stop there versus when they are likely to make the effort to consider other types of credibility as well, which might include seeking third-party endorsements or recommendations (i.e., reputed credibility) or checking the site's domain name suffix and internal/external links (i.e., presumed credibility), for example.

Some preliminary evidence points to the applicability of a dual processing model of credibility assessment. According to dual processing tenets outlined earlier, motivation and ability should be key factors in users' credibility assessment efforts. Flanagin and Metzger (2000) found that credibility assessment varied by information type, such that participants reported exerting significantly more effort to verify the credibility of reference information (defined as factual, non-news information such as self-help, health-related, or "how-to" information) than they did for entertainment information. To the extent that it is reasonable to assume Internet users will be more concerned about obtaining accurate reference as compared to entertainment information, this finding suggests that motivation indeed impacts credibility assessment behavior. Other studies have provided evidence that user ability also plays a role in credibility assessment. For example, some research has found that Internet experience and Internet self-efficacy are positively related to users' efforts to verify the credibility of information they find online (Flanagin & Metzger, 2000; Hong, 2006). Finally, other research has shown that users possessing different levels of motivation and ability pay attention to different criteria when judging the credibility of Web sites. Specifically, this work found that nonexperts and those who judge information on a Web page to be less personally salient tend to use simple heuristics such as visual appeal in assessing credibility more than do experts and people for whom information is personally salient and consequential. By contrast, experts and those who find online information to be salient were

more concerned about the quality of the site's information and source credentials when making credibility judgments (Flanagin & Metzger, 2007; Stanford, Tauber, Fogg, & Marable, 2002). While provocative, more research is clearly needed to fully understand the processes involved in Internet users' credibility determinations under different conditions, and therefore the skills users need to make those determinations.

Conclusion

This article summarizes much of what is known about the skills Internet users need to assess the credibility of online information. In addition, it describes and evaluates several models of online information evaluation and reviews the empirical research on how users themselves go about the process of determining the credibility of Internet information. Based on this, the article offers a number of recommendations for online credibility research, theory, and practice, and proposes alternative strategies to assist users in locating credible information on the Internet. Several overarching conclusions emerge from the review, including the fact that Internet users as a group are not willing to exert a great deal of effort in assessing the credibility of the information they find online, and that they place a premium on professional site design in their credibility appraisals.

Perhaps the most important conclusion from the review is that any effort to teach Web evaluation must be realistic in its expectations of users by recognizing that motivation is a key ingredient in users' willingness to undertake extensive effort to verify the credibility of information they find online. Educators could leverage this point by including information about the negative consequences of misinformation online, as a means to motivate users to critically evaluate information. Indeed, this is a fundamental theoretical tenet in other types of educational campaigns (see Rice & Atkin, 2001). As Fritch (2003) wrote:

We all recognise [sic] that some information is not important enough to require careful evaluation, but each individual must determine when this is true. And when necessary, users need the evaluative skills to make a proper determination about which information is trustworthy, on the Internet and in other venues. This is part of what it means to become a literate citizen in an information-rich, networked world. (p. 327)

In sum, the Internet has not so much changed what skills are needed to evaluate the credibility of information as it has changed the need for people to know when and how to exercise those skills.

References

- Alexander, J.E., & Tate, M.A. (1999). *Web wisdom: How to evaluate and create information quality on the Web*. Hillsdale, NJ: Erlbaum.
- Amsbary, J.H., & Powell, L. (2003). Factors influencing evaluations of web site information. *Psychological Reports*, 93(1), 191-198.

⁹For example, users may form judgments about the credibility of a Web site as a whole, the credibility of some information presented on a Web page, the credibility of the Web/Internet medium itself, the credibility of a particular e-mail message or blog, and so on. Metzger, Flanagin, Eyal, et al. (2003) presented a comprehensive discussion of this issue.

- Baertlein, L. (2005, May 6). Google trying to patent news ranking by quality. Retrieved July 24, 2007, from http://findarticles.com/p/articles/mi_zdext/is_200505/ai_n13636886
- Brandt, D.S. (1996). Evaluating information on the Internet. *Computers in Libraries*, 16, 44-46.
- Burbules, N.C. (1998). Rhetorics of the Web: Hyperreading and critical literacy. In I. Snyder (Ed.), *Page to screen: Taking literacy into the electronic era* (pp. 102-122). London: Routledge.
- Chaiken, S. (1980). Heuristic versus systematic processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752-766.
- Chaiken, S., & Trope, Y. (1999). Dual-process theories in social psychology. New York: Guilford Press.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73-96). New York: Guilford Press.
- Eysenbach, G. (2000). Consumer health informatics. *British Medical Journal*, 320, 1713-1716.
- Eysenbach, G., & Diepgen, T.L. (1998). Towards quality management of medical information on the internet: Evaluation, labeling, and filtering of information. *British Medical Journal*, 317, 1496-1500.
- Eysenbach, G., & Kohler, C. (2002, March 9). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal*, 324, 573-577.
- Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web: A systematic review. *Journal of the American Medical Association*, 287(20), 2691-2700.
- Fallows, D. (2005). Search engine users: Internet searchers are confident, satisfied and trusting—But they are also unaware and naïve. Report for the Pew Internet and American Life Project. Retrieved January 25, 2005, from http://www.pewinternet.org/PPF/r/146/report_display.asp
- Flanagin, A.J., & Metzger, M.J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly*, 77(3), 515-540.
- Flanagin, A.J., & Metzger, M.J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of Web-based information. *New Media & Society*, 9, 319-342.
- Fogg, B.J. (2003). *Persuasive technology: Using computers to change what we think and do*. San Francisco: Kaufmann.
- Fogg, B.J., Soohoo, C., Danielson, D.R., Marable, L., Stanford, J., & Trauber, E.R. (2003). How do users evaluate the credibility of Web sites? A study with over 2,500 participants. Paper presented at the Proceedings of the Conference on Designing for User Experiences, San Francisco. Available at: <http://portal.acm.org/citation.cfm?doid=997078.997097>
- Fox, B. (2005, April 30). Google searches for quality not quantity. *New Scientist*, 2497, 24. Retrieved May 12, 2005, from <http://www.newscientist.com/article.ns?id=mg18624975.900>
- Fox, S., & Rainie, L. (2002, May 22). Vital decisions: How Internet users decide what information to trust when they or their loved ones are sick. Washington, DC: Pew Internet & American Life Project. Retrieved June 5, 2005, from http://www.pewinternet.org/pdfs/PIP_Vital_Decisions_May2002.pdf
- Freeman, K.S., & Spyridakis, J.H. (2004). An examination of factors that affect the credibility of online health information. *Technical Communication*, 51(2), 239-263.
- Fritch, J.W. (2003). Heuristics, tools, and systems for evaluating Internet information: Helping users assess a tangled Web. *Online Information Review*, 27(5), 321-327.
- Fritch, J.W., & Cromwell, R.L. (2001). Evaluating Internet resources: Identity, affiliation, and cognitive authority in a networked world. *Journal of the American Society for Information Science and Technology*, 52(6), 499-507.
- Fritch, J.W., & Cromwell, R.L. (2002). Delving deeper into evaluation: Exploring cognitive authority on the Internet. *Reference Services Review*, 30(3), 242-254.
- Hong, T. (2006). The Internet and tobacco cessation: The roles of Internet self-efficacy and search task on the information-seeking process. *Journal of Computer-Mediated Communication*, 11(2), Article No. 8. Retrieved March 3, 2006, from <http://jcmc.indiana.edu/vol11/issue2/hong.html>
- Hovland, C.I., Janis, I.L., & Kelley, J.J. (1953). *Communication and persuasion*. New Haven, CT: Yale University Press.
- Johnson, T.J., & Kaye, B.K. (2000). Using is believing: The influence of reliance on the credibility of online political information among politically interested Internet users. *Journalism & Mass Communication Quarterly*, 77, 865-879.
- Johnson, T.J., & Kaye, B.K. (2002). Webbelievability: A path model examining how convenience and reliance predict online credibility. *Journalism & Mass Communication Quarterly*, 79, 619-642.
- Kapoun, J. (1998). Teaching undergrads WEB evaluation: A guide for library instruction. Retrieved July 24, 2007, from <http://www.ualberta.ca/~dmiall/Brazil/Kapoun.html>
- Kunst, H., Groot, D., Latthe, P.M., Latthe, M., & Khan, K.S. (2002, March 9). Accuracy of information on apparently credible websites: Survey of five common health topics. *British Medical Journal*, 324, 581-582.
- Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *Libraries and the Academy*, 4(3), 331-344.
- Metzger, M.J., Flanagin, A.J., Eyal, K., Lemus, D., & McCann, R. (2003). Bringing the concept of credibility into the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication Yearbook*, 27, 293-335.
- Metzger, M.J., Flanagin, A.J., & Zwarun, L. (2003). Student Internet use, perceptions of information credibility, and verification behavior. *Computers & Education*, 41, 271-290.
- Morahan-Martin, J.M. (2004). How Internet users find, evaluate, and use online health information: A cross-cultural review. *CyberPsychology & Behavior*, 7(5), 497-510.
- Morahan-Martin, J.M., & Anderson, C. (2000). Information and misinformation online: Recommendations for facilitating accurate mental health information retrieval and evaluation. *CyberPsychology & Behavior*, 3, 731-742.
- O'Keefe, D.J. (2002). *Persuasion: Theory and research* (2nd ed.). Thousand Oaks, CA: Sage.
- Petty, R.E., & Cacioppo, J.T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, IA: Brown.
- Rainie, L., & Hitlin, P. (2004). The use of online reputation and rating systems. Available at: http://www.pewinternet.org/PPF/r/140/report_display.asp
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of ACM CSCW Conference on Computer Supported Cooperative Work, Sharing Information and Creating Meaning* (pp. 175-186). ACM: Chapel Hill, NC.
- Rice, R.E., & Atkin, C.K. (Eds.). (2001). *Public communication campaigns* (3rd ed.). Thousand Oaks, CA: Sage.
- Rice, R.E., & Katz, J.E. (Eds.). (2001). *The Internet and health communication: Experiences and expectations*. Thousand Oaks, CA: Sage.
- Rice, R.E., & Katz, J.E. (2006). Internet use in physician practice and patient interaction. In M. Muneror & R.E. Rice (Eds.), *The Internet and health care: Theory, research and practice* (pp. 149-176). Mahwah, NJ: Erlbaum.
- Rice, R.E., Peterson, M., & Christine, R. (2001). A comparative features analysis of publicly accessible commercial and government health database Web sites. In R.E. Rice & J.E. Katz (Eds.), *The Internet and health communication: Experiences and expectations* (pp. 213-231). Thousand Oaks, CA: Sage.
- Rieh, S.Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, 53(2), 145-161.
- Rieh, S.Y., & Belkin, N.J. (1998). Understanding judgment of information quality and cognitive authority in the WWW. In C.M. Preston (Ed.), *Proceedings of the 61st ASIS annual meeting* (pp. 279-289). Silver Spring, MD: American Society for Information Science.

- Rosen, D.J. (1998, February). Driver education for the information super-highway: How adult learners and practitioners use the Internet [Series title: Literacy Leader Fellowship Program Reports, 2(2)]. Washington, DC: National Institute for Literacy.
- Scholz-Crane, A. (1998). Evaluating the future: A preliminary study of the process of how undergraduate students evaluate Web sources. *Reference Services Review*, 26(3/4), 53-60.
- Seidman, J.J. (2006). The mysterious maze of the World Wide Web: How can we guide consumers to high-quality health information on the Internet? In M. Murero & R.E. Rice (Eds.), *Internet and health care: Theory, research and practice* (pp. 195-212). Mahwah, NJ: Erlbaum.
- Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Smith, A.G. (1997). Testing the surf: Criteria for evaluating Internet information resources. *Public-Access Computer Systems Review*, 8. Retrieved July 24, 2007, from <http://epress.lib.uh.edu/pr/v8/n3/smit8n3.html>
- Stanford, J., Tauber, E.R., Fogg, B.J., & Marable, L. (2002). Experts vs. online consumers: A comparative credibility study of health and finance Web sites. *Consumer WebWatch Research Report*. Retrieved March 3, 2006, from <http://www.consumerwebwatch.org/dynamic/web-credibility-reports-experts-vs-online-abstract.cfm>
- Walther, C.N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 53(2), 134-144.
- Wilder, S. (2005, January 7). Information literacy makes all the wrong assumptions. *Chronicle of Higher Education*, 51(18), B13.

ARTICLE IN PRESS

Available online at www.sciencedirect.com

Information Processing and Management xxx (2007) xxx-xxx

**INFORMATION
PROCESSING
&
MANAGEMENT**
www.elsevier.com/locate/infoproman

Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context

Brian Hilligoss¹, Soo Young Rieh^{*}

School of Information, University of Michigan, 304 West Hall, 1085 South University Avenue, Ann Arbor, MI 48109-1107, United States

Received 18 April 2007; received in revised form 30 September 2007; accepted 8 October 2007

Abstract

This article presents a unifying framework of credibility assessment in which credibility is characterized across a variety of media and resources with respect to diverse information seeking goals and tasks. The initial data were collected through information-activity diaries over a 10-day period from 24 undergraduate students at three different colleges. Based on 245 information seeking activities from the diaries, the authors conducted individual interviews with participants and analyzed the transcripts using a grounded theory analysis. Three distinct levels of credibility judgments emerged: construct, heuristics, and interaction. The construct level pertains to how a person constructs, conceptualizes, or defines credibility. The heuristics level involves general rules of thumb used to make judgments of credibility applicable to a variety of situations. Finally, the interaction level refers to credibility judgments based on content, peripheral source cues, and peripheral information object cues. In addition, context emerged as the social, relational and dynamic frames surrounding the information seeker and providing boundaries of credibility judgments. The implications of the framework in terms of its theoretical contribution to credibility research and practices are discussed.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Information credibility; Credibility assessment; Information seeking behavior

1. Introduction

Every day people encounter more information than they can possibly use. Friends, colleagues, books, newspapers, television, and Web sites are just a few of the resources and media contributing to the flow of information. But all information is not necessarily of equal value. In many cases, certain information appears to be better, or more trustworthy, than other information. The challenge that most people then face is to judge which information is more credible. The concept of credibility has received considerable attention since the late 1990s when the Internet began providing a new information interaction environment that allowed users to seek for information and communicate with others in ways never before possible. As a consequence,

^{*} Corresponding author. Tel.: +1 734 647 8040; fax: +1 734 764 2475.

E-mail addresses: bhilligo@umich.edu (B. Hilligoss), rieh@umich.edu (S.Y. Rieh).

¹ Tel.: +1 734 764 1858; fax: +1 734 647 8045.

ARTICLE IN PRESS

2

B. Hilligoss, S. Y. Rieh / Information Processing and Management xxx (2007) xxx-xxx

researchers and practitioners in diverse fields including information science, marketing, management information systems, communication studies, human-computer interaction (HCI), and psychology have examined credibility assessment from a variety of different perspectives (Rieh & Danielson, 2007).

Many studies of credibility tend, however, to investigate credibility by relying on definitions, approaches, and presuppositions that are field-specific (Flanagin & Metzger, 2007). In information science, credibility has been understood as one of the criteria of relevance judgment used when making the decision to accept or reject retrieved information (Rieh & Danielson, 2007). Communication researchers have, on the other hand, been examining credibility as a research agenda distinguishing message credibility, source credibility, and media credibility (Metzger, Flanagin, Eyal, Lemus, & McCann, 2003). Management information systems (MIS) researchers have examined credibility issues in expert systems and decision support systems by querying people as to whether they believe the advice these systems provide. Consumer behavior researchers have addressed how consumers distinguish subjective and objective claims in e-commerce Web sites.

Consequently, previous studies of credibility have provided insights in relation to particular media such as the Web (e.g., Fogg et al., 2003; Huerta, 2003; Rieh, 2002), particular types of information such as political information (Johnson & Kaye, 1998, 2000), news information (Sundar, 1999), and scholarly information (Liu, 2004) as well as particular settings such as schools (Fidel et al., 1999). This approach is potentially problematic in that people may carry over strategies and criteria used for one type of information (e.g., political information) to another type of information (e.g., scholarly information) and vice versa. People also may rely on multiple types of media or resources in the course of a single information seeking episode. What is called for, then, is a framework that identifies common aspects of credibility assessment regardless of media, type of information, and environment of information use.

This paper proposes a unifying framework of credibility assessment in which credibility is characterized across various types of information resources and diverse information tasks including work- and school-related problems and personal interests. Importantly, this paper differentiates the terms credibility assessment and credibility judgment. Credibility assessment is herein seen as an iterative process involving one or more credibility judgments. As an information seeker encounters information, a series of judgments are made about the credibility of that information. These judgments are based on various factors to be discussed in greater depth in the following sections. Taken together, those judgments comprise the credibility assessment which feeds into the decision to accept or reject information.

This article begins with definitions of credibility and existing theoretical frameworks of credibility assessment which help to motivate the need for new research to develop a credibility assessment framework with respect to people's information seeking behaviors in which multiple information resources are used. The methods of data collection and analysis used in the present study are then presented. Results are discussed in the form of an emerging unifying framework that is grounded in empirical findings. Finally, implications and contributions of the framework are discussed with respect to the previous literature on credibility.

2. Review of related literature

2.1. Definition of credibility

Credibility has been defined as believability, trust, reliability, accuracy, fairness, objectivity, and dozens of other concepts and combination thereof (Self, 1996). It also has been defined in terms of characteristics of persuasive sources, characteristics of the message structure and content, and perceptions of media (Metzger et al., 2003). Some studies focus on the characteristics that make sources or information worthy of being believed, while others examine the characteristics that make sources or information likely to be believed (Flanagin & Metzger, 2007).

Despite the fact that communication researchers have been interested in source credibility since the 1950s, there is as yet no clear definition of credibility. The overarching view across definitions is believability. Credible people are believable people, and credible information is believable information (Tseng & Fogg, 1999). Most credibility researchers agree that there are at least two key dimensions of credibility: trustworthiness and expertise (Hovland, Janis, & Kelley, 1953), both of which contribute to the concept of credibility. Trustworthiness is a key factor in credibility assessment. A person is trustworthy for being honest, careful in choice

ARTICLE IN PRESS

B. Hilligoss, S.Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

3

of words, and disinclined to deceive (Wilson, 1983). Information is trustworthy when it appears to be reliable, unbiased, and fair. Expertise is "the perceived knowledge, skill, and experience of the source" (Fogg, 2003a, p. 124). Expertise is also an important factor because it is closely related to user perceptions of the ability of a source to provide information both accurate and valid. When people find that sources have expertise, they are likely to judge that information to be trustworthy. Assessment of sources' expertise come in multiple ways: people might have prior first-hand experience with a source; they might have heard about a source from other people; they might know that a source has a good reputation; and they might recognize expertise when a source has credentials, among others (Rieh, 2002). People's credibility judgments are subjective and implicit because they need to recognize expertise to conclude that the information is trustworthy.

Credibility differs from cognitive authority and information quality, while being closely related to these two concepts. Cognitive authorities are those people who actually exert influence on other people's thoughts by being recognized as proper (Wilson, 1983). According to Wilson, an authority's influence is thought proper because "he is thought credible, worthy of belief" (p. 15). Cognitive authorities are clearly among those who are considered to be credible sources. A person may be recognized as credible in an area even if not exerting influence on other people's thoughts. In Wilson's words, "Those we think credible constitute the potential pool of cognitive authorities on which we might draw" (Wilson, p. 16). Wilson claims that people do not attribute cognitive authority exclusively to individuals. Cognitive authority can be found in books, instruments, organizations, and institutions.

Information quality refers to people's subjective judgment of *goodness and usefulness of information* in certain information use settings with respect to their own expectations of information or in regard to other information available. Information quality is composed of five facets: usefulness, goodness, accuracy, currency, and importance (Rieh, 2002). These facets of information quality are not necessarily always consistent. Information may be accurate but not useful, useful but not important, important but no longer current, current but inaccurate, and so forth. In such cases, one of the questions people ask themselves would be whether they can believe what the information says or, if not, whether they can at least take it seriously (Wilson, 1983). Thus, credibility is a *chief aspect of information quality*.

2.2. Existing theoretical frameworks of credibility assessment

Several existing theoretical frameworks inform the understanding of credibility assessment. Fogg's (2003b) Prominence-Interpretation Theory grew out of a series of research projects on Web credibility conducted for four years at the Stanford Web Credibility Research Lab. This theory posits that two things need to happen for people to make a credibility assessment: the user notices something (prominence), and the user makes a judgment about what she/he notices (interpretation). Prominence is "the likelihood that a Web site element will be noticed or perceived" (p. 722). If the element is not noticed, it will have no impact on how the user assesses the credibility of the site. Fogg claims that at least five factors affect prominence: involvement (motivation and ability to scrutinize Web site content), topic of the Web site, task of the user, experience of the user, and individual differences of users. Interpretation, the second component of the theory, is "a person's judgment about an element under examination" (p. 723), or the evaluation of a Web site element in terms of being good or bad. According to Fogg's theory, various factors affect interpretation: user assumptions (culture, past experiences, heuristics, etc.), user skill/knowledge (level of competency in the site's subject matter), and context (user environment, user expectations, situational norms, etc.). This process of prominence and interpretation takes place more than once when a person evaluates a Web site because new aspects of the site are continually noticed and interpreted in the process of making overall assessments of credibility.

Rieh's (2002) model also looks at credibility assessment as an iterative process. While Fogg's (2003b) theory describes the credibility assessment process from the point when a user notices something in a Web site to the point when the user makes a judgment, Rieh's model begins earlier at the point at which a user makes a predictive judgment about which Web site will contain credible information and follows through to include evaluative judgments by which the user expresses preferences for information encountered. Rieh states that the kinds of factors influencing predictive judgments of information quality and cognitive authority differ from those that influence evaluative judgments. Her empirical study findings indicate that users tend to rely on their

ARTICLE IN PRESS

4

B. Hilligoss, S. Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

previous knowledge in terms of systems (system functions and structures) or topic area for making predictive judgments while their evaluative judgments are based on the characteristics of information objects (content, type of information object, and presentation). The characteristics of sources are consistently important criteria for both predictive judgments and evaluative judgments, according to Rieh's study.

Wathen and Burkell (2002) present a model of credibility assessment in which a staged process is laid out in the context of health information searching on the Web. They propose that a user's first task is to rate the credibility of the medium based on surface characteristics such as appearance/presentation, usability/interface design, and organization of information. The second task for the user is rating the source and the message. The evaluation of a source is often made in terms of its expertise/competence, trustworthiness, credentials, and so forth. The message is evaluated in terms of content, relevance, currency, accuracy, and tailoring. The third aspect of process involves the interaction of presentation and content with the user's cognitive state. Wathen and Burkell's model is also iterative given that the user makes judgments of the initial surface of Web sites and source cues via interactions with Web sites. Their model, however, has not been tested empirically.

Sundar (2007) presents four of what he calls "affordances" in digital media capable of cueing cognitive heuristics pertinent to credibility assessments: Modality (M), Agency (A), Interactivity (I), and Navigability (N). His MAIN model argues that these four technical affordances help to explain the perceived credibility of digital media and their offerings beyond what is explained by content characteristics. According to Sundar, the affordances have the power to amplify or diminish content effects on credibility because they can play the role of a moderator in a variety of psychological ways. Sundar further suggests that a deeper understanding of the role of heuristics in the credibility assessment process will clear up apparent contradictions in the credibility literature. While heuristics are not infallible and do not guarantee success, they likely appeal to many individuals who might cope with the deluge of information. Further discussion of the importance of heuristics in conceptualizing credibility assessment is found in Petty and Cacioppo's Elaboration Likelihood Model (ELM) to be discussed below.

2.3. Elaboration Likelihood Model (ELM)

The Elaboration Likelihood Model (ELM) provides a fairly general framework for understanding the basic process underlying persuasive communication, and as such it has often been adopted by credibility researchers to characterize credibility assessment processes and elements (e.g., Eastin, Yang, & Nathanson, 2006; Sundar, 2007; Wathen & Burkell, 2002). The ELM explains attitudinal changes in individuals as they encounter messages and the sources of those messages (Petty & Cacioppo, 1981, 1986). The two key constructs capable of affecting the amount and direction of attitude change are argument/message quality and peripheral cues. In the ELM, arguments are viewed "as bits of information contained in a communication that are relevant to a person's subjective determination of the true merits of an advocated position" (p. 133). Thus, one way to influence attitude in a persuasive message is to strengthen the quality of the arguments. Another way to influence attitude is to use a simple cue even in the absence of arguments. Peripheral cues can pertain to either source or message. An example of a peripheral source cue is the reputation of the source (highly respected author, etc.), while an example of a peripheral message cue would be the length of the message or the number of points made in the argument. The distinction Petty and Cacioppo make between content and peripheral cues may have important implications for credibility research given that "peripheral route" and "central route" characterize where people tend to put their effort and attention in evaluating information credibility when using digital media (Sundar, 2007).

Petty and Cacioppo (1981, 1986) also identify two types of message processing in which individuals engage when encountering messages: systematic and heuristic. Systematic processing is effortful and involves analyzing a message based on content cues. In many instances, however, individuals lack either the motivation or the ability to engage in systematic processing. At those times they resort to heuristic processing, basing their judgments of a message not on its content but on peripheral cues from the message and/or the source. Heuristics are practical rules or guidelines that aid in problem solving, decision making, and discovery (Newell, Shaw, & Simon, 1960), and as such tend to reduce mental effort (Petty & Cacioppo, 1986; Sundar, 2007).

Please cite this article in press as: Hilligoss, B., & Rieh, S. Y., Developing a unifying framework of credibility ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.10.001

ARTICLE IN PRESS

B. Hilligoss, S.Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

5

3. Methodology

The goal of this study was to better understand how people make credibility assessments in a wide variety of everyday life information seeking contexts. To accomplish this goal, undergraduate students were selected as potential participants under the assumption that their lives would involve a wide variety of information seeking activities across work, school, and personal pursuits.

3.1. Participants

A purposive sampling approach to recruit participants was used, ensuring that a variety of undergraduate students representing different sizes of colleges, academic majors, and genders were included. Recruiting was continued up to the point of theoretical saturation (i.e., when participants began repeating the same themes and the authors ceased learning anything new). Twenty-six undergraduate students were recruited from three different institutions in the US Midwestern state. The three institutions included a large research university, a medium-sized state university, and a community college. Two participants dropped out of the study before they could be interviewed, and their data are not reported here. So, in total, 24 students participated in the study. Initially, only first year students were recruited on the belief that they would likely be encountering a variety of novel information tasks in their new learning and living environments. At the research university, the sampling was limited to first year students; this limiting, however, proved extremely difficult at the other two institutions as recruitment for the study occurred primarily during the summer months when fewer first year students were on campus. Therefore, recruitment was extended to include undergraduate students at all levels.

Although the average age of the participants was 21.8 years, 11 participants were 18-year old. Half of the participants were in the first year of their undergraduate education. There were only three non-traditional students, including a 29-year old, a 31-year old, and a 43-year old. Fourteen were female, and ten were male. The participants' majors included engineering, dance, film, psychology, business, nursing, education, and pre-medicine, among others.

A pre-interview questionnaire was administered to gather background information on the participants. On average, participants said they used the Internet 3.6 hours each day and their institution's online library catalog system about 0.75 hour each week. Twenty-one participants (87.5%) reported they used email multiple times daily and went on the Internet for personal interests four or more times each week. Twenty participants (83.3%) also responded that they used the Internet for school research approximately four to seven days per week.

3.2. Data collection

Diaries have proven useful to researchers given that the method enables the collection of a wealth of detailed, actual information behaviors in specific contexts (Toms & Duff, 2002). Furthermore, diaries make it possible to capture various kinds of information activities such as topics in task, resources used, time spent on task, and outcome of search activity while they are still fresh in the mind of the participant and are useful for triggering memories of those details during a subsequent in-depth interview (Rieh, 2004). Thus, participants are not left to rely entirely on their memories to recall their information needs and activities, making diaries a useful means of capturing the salient aspects of real information seeking activities for discussion during later interviews (Cool & Belkin, 2002).

Participants were asked to record one information seeking activity per day for 10 days. The general instructions given to the participants were to: "Think about situations today in which you needed to find information. Choose the one that was most important to you; then answer the following questions, providing as much detail as you can." They were encouraged to include activities that involved the range of information sources (human, Web site, book, newspaper, etc.), because the intent was not to limit the investigation to any one system, source or medium. The diaries were kept online via a password-protected Web page. Participants answered a series of 11 specific questions about the activity to ensure important details were captured. The questions posed in the diary form included the topic of information seeking activity, goals of information

Please cite this article in press as: Hilligoss, B., & Rieh, S. Y., Developing a unifying framework of credibility ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.10.001

ARTICLE IN PRESS

seeking, familiarity with the topic, resources used, information seeking process, information use plans, and their next steps. No question directly related to credibility was included in the diaries. This is because questions about credibility in the diary may have changed the participants' behaviors and attitudes by having them focus on credibility issues during the process of information seeking.

Once the activities from the entire 10 days had been recorded, an in-depth interview was conducted. The researcher prepared two hard copies of the participant's diary and gave one to the participant so that both the researcher and the participant could read the entries in the diary during the interview. The interview was initiated by asking the participant to recount each activity, providing additional details about the topic, the resources used, and the step-by-step information seeking process. The participant was then queried about issues of credibility, including whether or not credibility was a concern and how it was judged. The participant was also asked to provide reasons for the judgments of credibility that they made. Additional questions in the interview included the participant's perceived levels of difficulty and confidence along with comments on the general seeking experience. The interviews were not, however, strictly limited to the 10 activities contained in the diaries: when appropriate, participants were asked to discuss comparable or contrasting activities with respect to their judgments of credibility generally.

3.3. Data analysis

The interviews ranged in length from 1 to 1.5 hours. The audiotapes of these interviews were transcribed and then analyzed using the grounded theory approach (Glaser & Strauss, 1967). The two authors conducted an initial content analysis separately on the same transcripts. The encodings of three transcripts resulted in the first draft of the coding scheme. Through discussions and explanations of how and why each code was applied, the coding scheme was revised. The three transcripts were finally recoded using the revised coding scheme, and the major themes emerged from the content analysis.

Two Library and Information Services masters students in the School of Information at the University of Michigan served as coders. They and the authors held weekly meetings from June to August 2006. The authors presented the coders with the coding scheme along with operational definitions and examples. Each coder was responsible for analyzing the transcripts of the twelve interviews, and in the first three weeks they coded two transcripts each week as well as brought questions and "uncodable" quotes to the meeting for the purposes of revising the coding scheme. Through the iteration of coding scheme revisions over several weeks, the coding scheme was finalized and the themes emerged more clearly. The coders entered their encodings using NVivo 2.0, qualitative data analysis software. The meetings of the group of four were continued until all the encodings were completed in order to resolve any differences in encodings between the two coders. The common themes and categories emerged in the grounded theory analyses were then broken down into seven major categories and 25 sub-categories as shown in Table 1.

This article focuses on the development of a unifying framework of credibility assessment. Thus, the three levels—construct, heuristics, and interaction—and the context of credibility judgments that emerged from the data analysis will be discussed in detail in the following chapters of this article. References can be made to

Table 1
Coding categories of data analysis

Major categories	Sub categories
Goal	Personal information need, academic achievement, entertainment, problem solving, and routines
Task	Subject searching, question answering, fact retrieval, known item searching, and advice/recommendations/opinions
Medium/Resource	Human resources, online, print, and mass media
Strategies	Starting point, use of multiple resources, and information seeking based on convenience and cost
Levels of credibility judgments	Construct, heuristics, and interaction
Context	Related to the selection of resources, applicability of judgments
Credibility judgment process	Prediction, evaluation, follow-up judgments (verification/re-evaluation/clarification/refusal)

Please cite this article in press as: Hilligoss, B., & Rieh, S. Y. . Developing a unifying framework of credibility ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.10.001

ARTICLE IN PRESS

B. Hilligoss, S.Y. Rieh / *Information Processing and Management xxx (2007) xxx–xxx*

7

other works by the authors (Rieh & Hilligoss, 2007; Rieh, Hilligoss, & Yang, 2007) for other themes such as goals, tasks, media/resources, credibility judgment process, and information seeking strategies.

4. A unifying framework of credibility assessment

A unifying framework of credibility assessment was developed as a result of data analysis. Three distinct levels of credibility judgments emerged: construct, heuristics, and interaction. The construct level pertains to how a person constructs, conceptualizes, or defines credibility. It is the most abstract level and as such involves broad notions of credibility that influence the person's judgments. The heuristics level involves general rules of thumb used to make judgments of credibility. This level is fairly general, being broad enough to apply to a variety of situations rather than specific to any particular situation. The interaction level refers to credibility judgments based on specific source or content cues. A summary of the three levels is presented in Table 2.

The three levels of credibility judgments above do not necessarily operate independently. In fact, as presented in Fig. 1, any or all of the levels interlink. This framework implies that each level affects the other levels in both directions from abstract to specific levels and vice versa. For instance, if a person constructs credibility in terms of reliability, that construct may influence the kind of heuristics that can help in identifying a resource likely to be reliable (e.g., official Web site). Heuristics may influence the ways in which a person assesses credibility by drawing attention to certain characteristics when interacting with the information resource. Sometimes a person makes a credibility judgment based on certain cues from a source of information and finds later that the judgment may contradict the original heuristic. In such cases, the current heuristic can be extended by adding a new general rule of thumb. On the other hand, once a person repeatedly relies on the same kind of cues relative to information objects within a single information seeking episode or across multiple information seeking episodes, the use of such cues can be established as personal heuristics. If a person's heuristics prove consistent over time, the heuristics may become the person's construct of credibility.

As Fig. 1 shows, context also emerged as an important factor influencing all three levels by playing a central role in the process of making a credibility judgment. The context is the social, relational, and dynamic frame of reference surrounding the person's information seeking process. In general, it creates boundaries around the information seeking activity or the credibility judgment itself. The context of credibility judgments can either guide the selection of resources or limit the applicability of such judgments.

Table 2
Three levels of credibility assessment

Level	Definition	Types	Influence
Construct	Conceptualizations of credibility	<ul style="list-style-type: none"> • Truthfulness • Believability • Trustworthiness • Objectivity • Reliability 	Provides a particular point of view for judging credibility
Heuristics	General rules of thumb that are broadly applicable to a variety of situations	<ul style="list-style-type: none"> • Media-related • Source-related • Endorsement-based • Aesthetics-based 	Provides useful ways of finding information conveniently and making credibility judgment quickly
Interaction	Specific attributes associated with particular information objects and sources for credibility judgments	<ul style="list-style-type: none"> • Content cues • Peripheral source cues • Peripheral information object cues 	Provides specific information source or object characteristics on which to base a judgment

Please cite this article in press as: Hilligoss, B., & Rieh, S. Y., Developing a unifying framework of credibility ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.10.001

ARTICLE IN PRESS

8

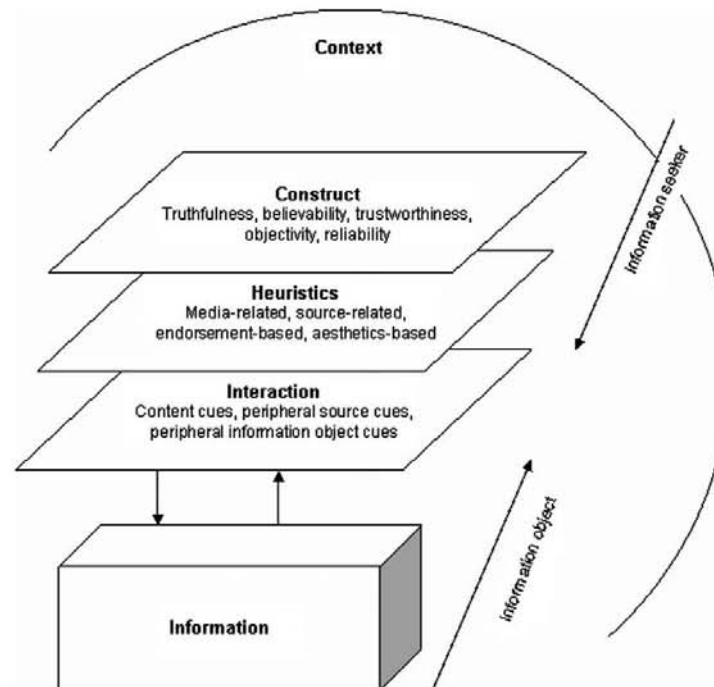
B. Hilligoss, S. Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

Fig. 1. A unifying framework of credibility assessment.

4.1. Construct level

The construct level is the highest in the framework because it concerns itself with how a person conceptualizes and defines credibility, providing a particular point of view for judging credibility in fundamental ways. The data analysis reveals that participants in this study conceptualize information credibility with respect to five different aspects: truthfulness, believability, trustworthiness, objectivity, and reliability. The distinct categories of a credibility construct indicate how a person defines credibility differently from other people at any given time.

There were a few cases in which judgments of credibility were made based on participants' estimations of the truthfulness of the information they encountered. "I guess what makes it credible would be: what type of information do they produce? Like, as long as they're producing the truth, and they're not altering it or providing false information, I would think of a source as credible" (P22 A11).² Other related definitions that participants used included accuracy, validity, correctness, and legitimacy.

Believability is another construct of credibility employed by several participants in this study. As one participant put it: "Credible is, like, believable" (P02 A11). In some cases participants applied this definition of credibility to the information itself, and in other cases they applied it to its sources. "...if somebody's not credible, then how can you believe in it?" (P13 A13).

Trustworthiness is the definition of credibility that participants mentioned most frequently. P16 said that "It comes down to being able to trust the source and trust the information" (P16 A12). As this quote demonstrates, the conceptualization of credibility as trustworthiness is one that can be applied to both information

² Throughout, quotations from the data are followed by a reference to the participant (P) and the information seeking activity (A) in which the quote occurred.

ARTICLE IN PRESS

B. Hilligoss, S.Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

9

and sources. In another example, when asked to talk about the credibility of an article on Wikipedia, one participant talked about the issue of trust by saying "But I do have a reason not to trust Wikipedia" (P19 A3). He went on to explain that "Because... anybody can edit any information on there... it's this thing where anybody can do anything to the information that's there. So, I just can't trust it because it could be modified" (P19 A3).

Another conceptualization of credibility that emerged during the interviews is objectivity. One participant defined credibility as "presenting [a] balance of facts and knowledge, showing different sides of the issues, not being biased, not being involved directly in what's going on is important" (P10 A2). This quote describes objective information ("showing different sides of the issues") and objective sources ("not being involved directly in what's going on").

Finally, several participants defined credibility in terms of reliability, a conceptualization of credibility closely linked with verifiability. One participant explained that "Being able to verify it with another source" makes for credible information (P18 A11). One interesting point that becomes apparent with this conceptualization of credibility is that it suggests certain strategies or behaviors for judging credibility. Participants who considered credibility as being about reliability felt they needed to seek out other sources of information or actually use the information in order to confirm its credibility.

Thus, it is evident that different people conceptualize credibility in different ways. This is not, however, to suggest that individuals have only one definition of credibility. In fact, the participants in this study often held multiple concepts of credibility. They applied certain constructs of credibility depending on the situation or type of information encountered, as shown in the following quote:

... I guess when I would trust information more than others kind of depends on how the Web site was set up. Like, for instance, the [amusement park] thing, I knew that was the Web site of the park, so I said, 'Okay, this is valid information,' or I assumed it to be. I guess they could be lying. When they say this is the official Web site of this then I kind of take it as, 'Okay, this is correct information.' But I guess... if you were trying to find out about hypertension or something, I would trust it more if... it said 'doctor' at the bottom of the page; MD at this hospital, or this foundation; like, the American something something I would trust it more. But probably the best place that I would believe information would probably be a doctor or some sort of clinic or medically trained person (P23 A11).

In this example the participant was concerned with whether or not the information told the truth (truthfulness) when discussing information related to an amusement park. When talking about health information, the participant constructed credibility in terms of trustworthiness.

4.2. Heuristics level

The heuristics level is comprised of general rules of thumb used to make decisions regarding credibility. These rules of thumb are fairly general, broad enough to apply to a variety of situations rather than being specific to any particular situation. Participants often explained their information seeking behaviors as doing what was "convenient" and "quick." The use of heuristics supports their objective of finding information quickly and conveniently. In many cases, heuristics allow individuals to almost instantly jump to a judgment of credibility without much substantial engagement with the information or source itself. As noted by Petty and Cacioppo (1981, 1986), systematic processing of the content of a message is time consuming and cognitively demanding, and in some cases individuals lack the motivation or ability to systematically process (or evaluate) the content of a message (or information object). In these cases individuals may engage in heuristic processing because they are either unwilling or unable to spend the time or effort required for systematic processing. The data herein revealed four categories of heuristics: media-related, source-related, endorsement-based, and aesthetics-based.

4.2.1. Media-related heuristics

Many of the heuristics that individuals hold pertain to specific media. Here, the term media is used broadly to refer to any media, format, or channel through which information is conveyed. In some instances, heuristics about different media are used to compare and contrast the relative credibility of those media. These general

ARTICLE IN PRESS

10

B. Hilligoss, S. Y. Rieh / Information Processing and Management xxx (2007) xxx-xxx

rules of thumb can exert various kinds of influence on the credibility assessment process, affecting such decisions as to where to look for information as well as the scrutiny or skepticism with which one can approach information found in a given media format.

Participants in the study expressed general heuristics about books, peer-reviewed journal articles, the Web, blogs, and libraries. Books and scholarly journal articles were consistently perceived as credible media, particularly in comparison with the Internet as shown in the comment that "People can just make up sites and stuff online, but with a book you have to go through publishing companies and getting that whole book [publishing] process. It just goes through so many procedures to make sure that it's credible, but the Internet, it's so easy [to post there]" (P02 A10). Knowledge of the peer-review process also lies behind the positive heuristics of scholarly journal articles. One student said, "I know it's really hard to become published, so, I'm assuming that it's credible" (P06 A8). Similarly, another participant explained that "With a peer reviewed article, I know that the person saying it is supposed to know something about the topic. ... I know that other people who are supposed to know something of the topic agreed with them or at least consider what they're saying a reliable possibility" (P01 A12). Thus, a general, if not detailed, understanding of the peer-review publishing process lay behind the confidence participants expressed in the credibility of books and scholarly journals.

Many participants are well aware of the notion that "... anybody can write anything on the Web" (P20 A7) and so tend to evaluate more carefully the information on the Web. Teaching in schools seems to play a part in the participants' heuristics about the Web as indicated in the statement that "... in high school I pretty much learned that everything you read [on the Web] is not true, and you can't trust everything that you read" (P22 A6). The concerns about the Web-based information did not, however, prevent the participants from using the information found on the Web. Rather, the heuristics about specific media will at times increase or decrease concerns about credibility, which can influence the scrutiny an individual uses in evaluation.

Some participants differentiated between types of Web sites, comparing and contrasting blogs to other Web sites. Participants who spoke of blogs perceived them as opinion-based and thus lacking credibility. As one participant put it, "I would say a blog would have a lot less credibility than a professional Web site; a lot of times they don't have first hand information" (P09 A5). That same participant elaborated to the effect that "[On] a blog, someone could say anything they want and just keep going. There's no journalistic integrity required or anything, because there's no repercussions" (P09 A5). Other participants noted that blogs can be useful as "spring boards," providing new ideas and leads for information seeking. One participant explained "So, I might use the information I found in a forum or a blog to point me in a direction to find out if something is true, but I wouldn't take that information as credible" (P10 A5). Thus, while people may perceive certain media or specific media formats to be non-credible, they still may have good reason for using those media.

4.2.2. Source-related heuristics

Analysis of the data revealed the participants employ heuristics about sources of information in two ways: familiar vs. unfamiliar sources and primary vs. secondary sources.

In terms of familiarity, the most common heuristics mentioned by the participants is that familiar, known sources are more credible than unfamiliar ones. One participant reported that "If it's from a trusted source. ... I would probably trust [it] more than, you know, when you ask someone random for directions; you don't really know if what they are saying is right" (P05 A11). Use of the word "random" is notable. Another participant spoke about random Web sites (P01 A11), while yet another participant referred to random organizations or random companies (P15 A2). In these cases, the heuristic is that unknown or random sources of information are less likely to be credible than known ones.

The second source-related heuristics is related to the dichotomy of primary vs. secondary sources. In general, participants perceived primary sources as being credible, with the "official site" being perceived as most credible. One participant stated that "I would say information is more credible, more trustworthy ... if the source is sort of the official source. ... regarding what you're looking for" (P08 A12). Another participant's characterization of the heuristics of secondary sources held that "... if it's like a second[ary] source, like, a paper someone has written online or a blog or an article even, then I kind of second guess whether it's a true Web site, whether that information is valid. So it depends on the site I'm on" (P03 A11).

ARTICLE IN PRESS

B. Hilligoss, S.Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

11

4.2.3. Endorsement-based heuristics

A third kind of source-related heuristics involves endorsements. Here, participants perceived a particular information object or source to be credible because it has been endorsed, recommended, or otherwise upheld by knowledgeable and trusted individuals. Those individuals might be known (friends, family, colleagues, etc.), or not known (e.g., experts in a particular field). Endorsements may also come in the form of citations, one participant explaining that "If [the book] had been cited in other history papers, then I know that people find it reliable" (P07 A8).

Organizations are also seen as givers of endorsements. Some participants felt that if the Web site of a trusted organization provided a link to another Web site, then that trusted organization was effectively endorsing the other Web site. One participant explained that "...if it's something that [the University] will send a link to, or a Web site I have used sends a link to, I trust that more. It's more credible to me because if they're associated with it, then they've done their checks" (P12 A12).

The findings demonstrate that some participants perceive popularity as a form of endorsement, following the general rule of thumb that information sources and objects widely used are more likely to be credible. One participant talked about his trust in several Web sites that review information technologies, calling them "reputable" and "respected in the industry." When asked to explain how he knew these Web sites were well respected, he replied that "...the fact that they're so popular, it's not like a perfect indicator, but it usually indicates they have some reputability" (P09 A5). The perception of popularity as a measure of credibility is not, however, restricted to Web sites. Another participant explaining why he might trust a moving company more if it were popular said "I think my general conclusion is that [if] everyone's using them. ..., then they must be a good company and a trusted company" (P15 A2). Another participant talked about the theory of global warming and how his own opinion about that theory had changed over time as he encountered "more and more and more" scientific articles supporting the theory. He said he now believed the theory because of "The overwhelming difference in proportion and amount [of scientists on the two sides]. It was like one hundred to one, really" (P16 A11). Nonetheless, he acknowledged that "The majority doesn't always mean it's correct" (P16 A11).

4.2.4. Aesthetics-based heuristics

A number of participants used heuristics that connected credibility and aesthetic design in Web sites. As one participant expressed it, "I also judge Web sites based on the quality of their layout. It's kind of snobbish, but if something looks like it's been put together by a five year old on his first GeoCities page, like, lots of blinky things and whatever, then I usually think the Web site's probably crap, because they haven't put a whole lot of effort into it" (P18 A8). One participant offered that "I guess a professional design of a Web site, like a nice and clean layout kind of—generally, if someone puts that much time into designing the Web site, they put more time into what they're saying. It's not always true, but it's kind of a good rule of thumb" (P09 A10).

4.3. Interaction level

The third and final level of influence is the interaction level, in which specific attributes associated with particular information objects and sources enter into an individual's judgment of credibility. In contrast to heuristics, which are broad and widely applicable, interaction level judgments are unique to a specific information object or source that participants encounter. The data analysis reveals that the interaction level involves three types of interactions: those with content cues, those with source peripheral cues, and those with information object peripheral cues.

4.3.1. Interactions with content cues

Content refers to the message itself, that is, the substantive information being conveyed. Interactions with content are arguably the most cognitively demanding of the three kinds of interactions (Petty & Cacioppo, 1986). This study finds that the use of personal knowledge to evaluate information is the primary method by which people interact with content from a credibility assessment perspective. P20, researching a particular Islamic scholar for a school assignment, conducted a Google search and found an article about him on

ARTICLE IN PRESS

Wikipedia. Since she already knew something about the scholar, she was able to recognize several errors in the content provided on the Wikipedia site. The following quote explains her interactions with the content with respect to her prior knowledge:

It's usually like, for example, on that one site that I went to on Wikipedia for that one article that I knew wasn't credible. I knew it wasn't credible because of some of the small mistakes in there. And others were pretty big mistakes that directly contradicted what I knew about the person who I was looking up. And so in that sense I knew it wasn't credible (P20 A11).

When the participants lacked the knowledge required for judging the credibility of information, they took other strategies for making credibility judgments based on content cues. For instance, participants looked for multiple sources with concurring information in order to verify content. One participant, researching the effectiveness of evidence-based substance abuse prevention programs, explored numerous scholarly articles. He explained why he felt that a particular article he found was credible: "All of the information follows suit with what others studies have said. It kind of falls in line with that" (P16 A7). Similarly, another participant was looking for the date for a particular upcoming community event. She conducted a search on Google, clicking on several of the links in the results set even though the first one she tried provided the date. She explained her reason for doing this as "Just reassurance—because I don't know which one is like—some sites on the Internet are not credible. So I wanted to make sure that it was the right information, and by giving the same date a couple of times, I believed it" (P22 A6).

4.3.2. Interactions with peripheral source cues

Any cue pertaining to a source that was used in a judgment of the credibility of information was deemed a peripheral source cue. In this study, individuals, groups, organizations, and any other social aggregate were considered sources. Source peripheral cues included affiliation, reputation, author's educational background, type of institution, and so forth.

Many participants explained their confidence in the credibility of information they received from friends, family, and other acquaintances. For example, one participant trusted her roommate for help on a sociology assignment because she knew the roommate was also taking the class and "doing pretty well" (P05 A1). Another participant trusted his friend for advice on ice skating because he knew "she has been skating many years—10 years" (P07 A2). Still another participant accepted input from his track coach about how best to treat an injury because he knew the coach "has been running marathon for nearly thirty years" (P15 A3). A different participant knew she could ask her boyfriend for information on just about any topic and have confidence that he would be knowledgeable because "he's always doing research on everything" (P06 A4). These are just a few examples of how knowledge about the habits, behaviors and talents of another person can influence a person's credibility judgment.

Past experience with sources sometimes quite significantly influenced a participant's subsequent judgments. Several participants explained their selection of Web sites by pointing to positive prior experiences with those Web sites. One participant was looking for information about an upcoming performance by a musical band. He checked two Web sites, both of which he had used previously, and found the information he needed. When asked whether he felt the information might have been not credible, he replied, "Not really. Both the sites I've been to many times before and completely trust...so I didn't have any issues like that" (P04 A2). Previous experiences can also boost a person's confidence in an e-commerce Web site. One participant explained why she used Amazon.com saying, "I've used them in the past. I've ordered things from them. I haven't had any bad experiences with ordering things or pre-ordering items" (P12 A8). First-hand knowledge can also make people doubt the credibility of information sources. One participant asked his friend for information about an upcoming football game but then decided to double check the information with another, more knowledgeable friend. In explaining why he was hesitant to trust the first friend he said "...he has a history of just sort of saying whatever when I ask him these trivial questions, because it's not a big issue. He doesn't really consider it a big issue, so he'll just say whatever" (P08 A10).

In addition to their personal experiences, people may also be influenced by the experiences of others when making credibility judgments. One participant trusted a book she called "The Gardener's Bible," explaining "I do remember hearing from other people a long time back that they'd used it and found it helpful. And so I

ARTICLE IN PRESS

B. Hilligoss, S.Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

13

thought that's probably a good bet" (P20 A3). A different participant talked about an activity in which she needed more information about a female musical artist. She explained why she turned to and believed the information she found on a particular Web site by citing her stepfather's experience with that Web site:

My step-dad actually had a band, and they had a CD on there, and [the Web site] asked [the band] to write the biography of it because they [the Web site] are kind of small, and it's a lot cheaper way to do it. So, I figured it was probably, if not [the musical artist] who wrote it, someone like her manager or someone who was familiar with her (P14 A4).

As these examples illustrate, knowledge of a source, whether first or second-hand, influenced credibility judgments as participants interacted with sources.

4.3.3. Interactions with peripheral information object cues

The third kind of interaction involves peripheral cues derived from the information object. Most commonly these cues pertain to the appearance or presentation of the information. Considerations of presentation tend to involve the external, aesthetic aspects of the information object, or the emotional effect of interaction with the object. As noted above, many participants used aesthetics-based heuristics to judge information credibility. When applying such heuristics to a specific information object, such as examining how well a particular Web site is designed, an individual uses information object peripheral cues to judge credibility on the interaction level. One participant's effort to contrast Google and Yahoo! is illustrative:

Google seems to be more legitimate simply because there's no advertisement and all they have is the search [box] on their Web site. If you go to Google.com it says "Google," the search box, and then you can search ... different things. Yahoo! has all this stuff: news, weather, advertisements, links to movies and stuff. So, Google appears to be more credible because it doesn't have any of the other stuff. It has that appearance that we are here to find information to help you search whatever information, while Yahoo! seems to be more kind of for entertainment or different kinds of things (P10 A5).

In addition to cues pertaining to the appearance or presentation of information, participants also considered cues derived from the language used. Referring to articles he had read on zdnet.com, cnet.com, and slashdot.com, one participant said, "I think they're well written in general, so I guess that leads it to be more reputable" (P9 A4). In another example, P10 looked online for a listing of Martin Luther's Ninety-five Theses. He explained why he trusted the listing he found. "Well, I mean it was the old English. The wording was kind of complex unnecessarily. It sounded like something from that period" (P10 A9).

Information object peripheral cues can also evoke affective responses that may influence credibility judgments. In talking about why he judged a particular Web site about Albert Einstein to be credible, one participant said there was a "scientific mood to the Web site. It's hard to describe" (P8 A5). Others talked about information "seeming" or "feeling" correct or credible but were often unable to elaborate. Such judgments, based on interactions with information objects, are shaped by peripheral cues pertaining to those objects.

4.4. Context of credibility judgments

The data analysis reveals that contextual factors can intervene and influence credibility judgments by constraining selection of resources for a particular information seeking activity. For instance, P08 talked about a paper he was writing for a class assignment. While he was comfortable using the Web and felt that the Web provided many resources of high quality and credibility, he noted that he was reluctant to use many Web resources out of fear that they would be perceived by his professor as non-credible simply because they were Web-based. He elaborated by stating that "If it's something that I could find an actual book on, I like that too, because just using the Internet—I'm pretty nervous about just using the Internet—just because professors usually don't like that, and it's good to include some actual books" (P08 A3). In this example, the participant followed a strategy for selecting resources that was based not entirely on his own judgment of what constituted credible information but partly as well on the school context to accommodate a professor's judgment criteria.

Participants sometimes accepted certain resources as credible primarily by relying on the context in which they encountered the resources. For example, one participant explained her confidence in the textbook she was

ARTICLE IN PRESS

using for one of her classes with "...the same guy who wrote the textbook wrote my test. So, I would trust it" (P05 A8). In this example one particular resource, the textbook, was intricately intertwined with the context out of which her information need had arisen. A similar example was found with a participant who turned to the professor to clarify the class assignment. She said "... I knew I could trust my professor to give me the correct information because it was his problem I was working on" (P01 A12). Again, selection of resources and judgments of credibility were guided by the context in which the information need emerged.

Another finding from the data analysis is that contextual factors also enter into judgments of credibility by bounding or otherwise limiting the applicability of those judgments. Several instances were found in which contextual factors both led the participants to make certain credibility judgments and prevented them from confidently extending those judgments to other contexts. These cases were most frequently noted when participants made credibility judgments in the context of their classes. Several participants talked about the credibility of the information they received from their class instructors, acknowledging that while the information might not be credible in the world outside the class, it was credible within. As one student expressed it, "It's reliable in the context of the class. It may or may not be reliable in the real world" (P01 A12). Another said, "I would assume it's credible, but I also know it's completely credible within the class because the class is designed around it. So, even if it's false, it's true within the bounds of the classroom" (P09 A3).

Context not only bounds credibility judgments of information received from instructors but also bounds judgments of other sources as well. For instance, one participant commented on the information found in her course textbook. "Well, it isn't necessarily that the book would be right. It's that everything was taken from the book that would be on the exam. So, if the book wasn't right, like, technically, then, it was irrelevant. It'd be irrelevant to my goal [of passing the test]" (P06 A5). Comments like these indicate that credibility is not always viewed as absolute but rather may be seen as relative given its relationship to certain contexts of information seeking and use.

4.5. Interplay among construct, heuristics, interaction, and context

As mentioned briefly at the beginning of Section 4, the multiple levels and contextual factors are interlinked in credibility assessment. This section provides examples from the data which illustrate how multiple levels and contextual factors can shape and influence credibility judgments.

One of the common credibility judgment strategies that the participants took was to apply their own heuristics of credibility assessment to a certain information seeking activity. In many cases, it proved to be useful. In one example, a participant who was planning to move from the United States to Australia wanted to find out whether or not she would be able to work legally in Australia. When beginning her search for this information, she thought of the Australian embassy and looked for its Web site. She explained the reasoning behind her actions this way: "... it's sort of the official source, I guess, of information for people who are immigrating to Australia. I turned to it as a credible source, I guess, because it's the official line" (P24 A5). Embedded in her explanation is the very common heuristic that official sources are credible sources. Ultimately, her heuristic proved useful because she found the information she needed on the embassy's Web site. In this example, the participant used heuristics in the process of information seeking to help her identify a resource that was likely to be credible. As she interacted with that resource, nothing about it contradicted that prediction of credibility derived from the heuristic. Along this same line, another participant explained why he trusted information he found in the textbook for his Spanish class: "I didn't have any reason not to trust it" (P10 A1). Thus, as long as the participants do not encounter any cues on the interaction level that contradict or challenge their heuristics, they appeared to be fairly confident in their credibility judgments based solely on heuristics.

Heuristics are sometimes formed around the roles that a source plays in the world. Librarian is one such role which several participants perceived as a credible source of information. One participant's search for fiction at her college library illustrates how this heuristic played out in relation to the interaction level. Not knowing whether or not the library collection included popular fiction or how to find it if it did, the participant decided to seek help from the librarian. The participant explained: "Because she [librarian] works here so I figured she knew how the system worked and where to find it. I didn't think she would direct me in the wrong

ARTICLE IN PRESS

B. Hilligoss, S.Y. Rieh / *Information Processing and Management xxx (2007) xxx–xxx*

15

area to find something” (P12 A1). This example indicates that nothing about the interaction with the librarian contradicted the heuristic, so the participant was able to keep her confidence in the librarian’s credibility.

While heuristics are useful, they have their limits. When factors from other levels intervened, participants deviated from heuristics. For example, one participant expressed heuristics that educational and non-profit Web sites are more credible than commercial ones. As he put it, “.coms ‘might not be as reliable as the other types, like .edu and .orgs. . .’” (P11 A4). While this general rule of thumb might have been useful in guiding him in previous information seeking activities, he made an exception in the case of searching for information about bi-polar disorder for a paper he was writing. He turned to WebMD, a commercial health information Web site saying, “. . .when you’re doing a research paper, try to stay away from the.coms because [they] might not be as valid, but I just also assumed that WebMD would be pretty reliable” (P11 A4). In this example, interactions with peripheral source cues on a commercial Web site influenced the participant to go against the heuristic and to judge the WebMD site credible.

Another example from this study indicated how context could prompt participants to make exceptions to the heuristics. As mentioned earlier, many participants ascribed credibility to textbooks assigned by professors. This heuristic was applied directly in two different activities of one participant as she discussed the textbooks for her anthropology and astronomy classes. However, the context surrounding her art history class, made her less trustful of the textbook:

[In] my art history class, my professor disagrees with the text book quite a bit, and there was this one theory about art after the Black Death, like, saying how it changed. And that’s, like, our textbook went along with this certain theory of art history. I don’t remember the name, but [the professor] was completely against it. She said, “Don’t read this. It’s not right.” So, for that class I know at least not to look at the book for everything. You have to go to lecture (P05 A9).

This participant did not fully distrust the textbook; however she did not accept it without question as she did for her other courses. She said that she needed to attend the lectures to know what information in the textbook she should believe and what she should not. Furthermore, her caution stemmed not from any cues on the interaction level. That is, nothing about content, source, or book itself made her concerned about the credibility of information. Instead, the context in which the textbook was used influenced her credibility judgment.

5. Discussion

The purpose of this research is to develop a theoretical framework of credibility assessment in which people’s credibility judgments are extended across multiple media for a variety of information seeking goals and tasks. Data from 24 in-depth interviews with college students seeking information through a variety of media and channels for work, school, and personal interests were collected and analyzed in order to develop a framework of credibility assessment. This framework involves the three levels of credibility judgments: construct, heuristics, and interactions, any or all of which may be involved in an attempt to assess the credibility of information encountered. Context emerged as an important factor that influences the three levels. This framework also indicates that these multiple levels interplay with each other in shaping a person’s credibility judgments.

Previous credibility studies have tended to use the measures of credibility concept by presenting multiple terms with which to study participants when examining credibility issues. To evaluate the credibility of newspapers, TV news, and online news, researchers have used measures such as fairness, accuracy, factuality, truthfulness, believability, completeness, precision, objectivity, lack of bias, trustworthiness, objectivity, depth, informativeness (Bucy, 2003; Johnson & Kaye, 2000; Meyer, 1988; Newhagen & Nass, 1989; Salwen, 1987; Sundar, 1999). It is, however, unclear whether these measures are predictors of perceived credibility or are underlying dimensions of the credibility concept itself (Newhagen & Nass, 1989). Multiple constructs of credibility that characterize conceptualizations or definitions of credibility were identified: truthfulness, believability, trustworthiness, objectivity, and reliability. At least two interesting findings were noted. First, each person seeming to possess his/her strong sense of what credibility means. Second, a person may have multiple constructs of credibility, with the construct being closely related to the type of information to be evaluated. For instance, when the person evaluates news information, objectivity could be a more important construct than others. When evaluating information from primary sources, reliability may play an important role for

ARTICLE IN PRESS

credibility constructs. With health information, for instance, the question of whether people can trust the source of information would be more important than any other constructs. The framework herein indicates that linking the constructs of credibility with types of information will be a useful direction for future research in credibility.

A few researchers (e.g., Flanagin & Metzger, 2000; Sundar, 2007), have paid attention to the heuristics of credibility assessment lying behind general credibility perceptions. The present study contributes to the literature of credibility by identifying the role that heuristics plays in credibility judgments. As general rules of thumb or principles for evaluating information, heuristics help people to judge information credibility somewhat intuitively while retaining some consistencies across different information seeking and use situations. For example, as people develop heuristics about certain digital media, they may not need to examine various characteristics of information objects or sources at the interaction level in order to evaluate the information they encounter. Similarly, if people hold heuristics about the aesthetic aspects of Web sites, they tend to apply these heuristics across various Web sites.

Most previous credibility research has focused on cues on the interaction level (e.g., Fogg et al., 2001, 2003; Hong, 2006; Rieh, 2002). The results of this study show that credibility judgments taking place when interacting with specific cues are affected by credibility judgments made on the other two levels. For instance, a person may have heuristics that the information issuing from commercial Web sites is not credible compared to that issuing from government or educational Web sites. However, when the content is read carefully, the person may trust the information from commercial sites as sounding credible. Once the person gains this experience, it would become a heuristic that may be used in future credibility assessments. The findings also imply that people's perceptions of context influence credibility judgments on the interaction level in fundamental ways that go beyond the appearance, source, and content of information.

This study's findings reveal that credibility is not viewed as an absolute attribute of information or sources. Rather, credibility is seen as relative to the social context in which information seeking is pursued and credibility judgments are made (Rieh & Hilligoss, 2007). Awareness of the boundaries of credibility assessment helps participants to determine their selection of resources by filtering out stimuli in the information seeking process. The contextual factors also influence credibility assessment by "bounding" or otherwise limiting the information use environments.

6. Conclusion

Given that this framework was developed from a qualitative study of 24 participants, the findings need to be discussed with caution. Nonetheless, the unifying framework provides a fruitful initial base with which to understand better the multiple-layered and complex nature of credibility assessment for future research on credibility. This article contributes to the understanding of credibility in three distinct ways.

First, credibility judgments can be understood by taking into consideration multiple levels from construct to heuristics and interaction that go beyond focusing on cues for assessment. The cues that people rely on to make credibility judgments have been the focus of much credibility research because these cues are easily observed by researchers and discussed by study participants. However, the framework herein suggests that credibility assessment needs to be understood beyond the level of interaction, by incorporating much broader perspectives given that people also make credibility judgments in their process of developing constructs, relying on heuristics, and considering contexts.

Secondly, this research has implications for methodology when investigating credibility. Credibility assessments are often made internally, and it might prove difficult for many people to articulate their cognitively processed judgments. The methodology used in this research took a naturalistic approach in which participants were asked to keep track of their information seeking activities in diaries. In the interviews, questions were primarily asked about information seeking resources selected and strategies taken. That is because credibility judgments are embedded in the information seeking process as people tend to judge the value of credibility implicitly when deciding where to find information and what to select.

Third, the framework herein demonstrates the importance of understanding credibility across multiple media and information resources. The findings indicate that participants relied on multiple types of media and resources within one information seeking episode. More importantly, they often preferred to make cred-

ARTICLE IN PRESS

B. Hilligoss, S.Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

17

ibility judgments by comparing the information retrieved from different resources. Thus, rather than focusing on credibility issues on the Web, in online news, or in electronic scholarly journals, for example, credibility should be investigated by taking into consideration multiple kinds of information media and resources that people are likely to use for their information seeking processes.

Acknowledgements

The authors are grateful to Jiyeon Yang and Aiko Takazawa for providing research assistance in coding the interview transcripts using NVivo 2.0. Thanks also to Professor Joan Durrance and the members of the Information Behavior Theory Research Group at the School of Information for their insightful suggestions and helpful discussions on an early draft. The authors also thank the anonymous reviewers for constructive comments.

References

- Bucy, E. P. (2003). Media credibility reconsidered: Synergy effects between on-air and online news. *Journalism and Mass Communication Quarterly*, 80(2), 247-264.
- Cool, C., & Belkin, N. J. (2002). A classification of interactions with information. In H. Bruce, R. Fidel, P. Ingwersen, & P. Vakkari (Eds.), *Emerging frameworks and methods: Proceedings of the fourth international conference of library and information science (CoLIS4)* (pp. 1-15). Greenwood Village, CO: Libraries Unlimited.
- Eastin, M., Yang, M.-S., & Nathanson, A. I. (2006). Children of the Net: An empirical exploration into the evaluation of internet content. *Journal of Broadcasting and Electronic Media*, 50(2), 211-230.
- Fidel, R., Davies, R. K., Douglass, M. H., Holder, J. K., Hopkins, C. J., Kushner, E. J., et al. (1999). A visit to the information mall: Web searching behavior of high school students. *Journal of the American Society for Information Science*, 50(1), 24-37.
- Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of internet information credibility. *Journalism and Mass Communication Quarterly*, 77(3), 515-540.
- Flanagin, A. J., & Metzger, M. J. (2007). Digital media and youth: Unparalleled opportunity and unprecedented responsibility. In M. Metzger & A. Flanagin (Eds.), *Digital media, youth, and credibility. MacArthur foundation series on digital media and learning* (pp. 5-28). Cambridge, MA: The MIT Press.
- Fogg, B. J. (2003a). *Persuasive technology: Using computers to change what we think and do*. San Francisco: Morgan Kaufmann.
- Fogg, B. J. (2003b). Prominence-Interpretation Theory: Explaining how people assess credibility online. *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, 722-723.
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., et al. (2001). What makes Web sites credible? A report on a large quantitative study. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 61-68.
- Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How do users evaluate the credibility of Web sites? A study with over 2500 participants. In *Proceedings of the 2003 conference on designing for user experiences (DUX'03)*. <http://portal.acm.org/citation.cfm?id=997097&coll=ACM&dl=ACM&CFID=36236037&CFTOKEN=18606069>.
- Glaser, B. G., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.
- Hong, T. (2006). The influence of structural and message features on Web site credibility. *Journal of the American Society for Information Science and Technology*, 57(1), 114-127.
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. New Haven: Yale University Press.
- Huerta, E. (2003). *The credibility of online information*. Unpublished doctoral dissertation. Claremont Graduate University.
- Johnson, T. J., & Kaye, B. K. (1998). Cruising is believing? Comparing internet and traditional sources on media credibility measures. *Journalism and Mass Communication Quarterly*, 75(3), 619-642.
- Johnson, T. J., & Kaye, B. K. (2000). Using is believing: The influence of reliance on the credibility of online political information among politically interested internet users. *Journalism and Mass Communication Quarterly*, 77(4), 865-879.
- Liu, Z. (2004). Perceptions of credibility of scholarly information on the Web. *Information Processing and Management*, 40, 1027-1038.
- Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. In P. J. Kalbfleisch (Ed.), *Communication yearbook* (Vol. 27, pp. 293-335). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Meyer, P. (1988). Defining and measuring credibility of newspapers: Developing an index. *Journalism Quarterly*, 65, 567-574.
- Newell, A., Shaw, J. C., & Simon, H. A. (1960). *Report on a general problem solving program*. Paper presented at the international conference on information processing, Paris.
- Newhagen, J., & Nass, C. (1989). Differential criteria for evaluating credibility of newspapers and TV News. *Journalism Quarterly*, 66, 277-284.
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitude and persuasion: Classic and contemporary approaches*. Boulder, CO: Westview Press.
- Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. *Advances in Experimental Social Psychology*, 19, 123-205.

Please cite this article in press as: Hilligoss, B., & Rieh, S. Y., Developing a unifying framework of credibility ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.10.001

ARTICLE IN PRESS

18

B. Hilligoss, S. Y. Rieh / *Information Processing and Management xxx (2007) xxx-xxx*

- Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society of Information Science and Technology*, 53(2), 145-161.
- Rieh, S. Y. (2004). On the Web at home: Information seeking and Web searching in the home environment. *Journal of the American Society for Information Science*, 55(8), 743-753.
- Rieh, S. Y., & Danielson, D. (2007). Credibility: A multidisciplinary framework. In B. Cronin (Ed.), *Annual review of information science and technology* (Vol. 41, pp. 307-364). Medford, NJ: Information Today.
- Rieh, S. Y., & Hilligoss, B. (2007). College students' credibility judgments in the information seeking process. In M. Metzger & A. Flanagin (Eds.), *Digital media, youth, and credibility. MacArthur Foundation Series on Digital Media and Learning* (pp. 49-72). Cambridge, MA: The MIT Press.
- Rieh, S. Y., Hilligoss, B., & Yang, J. (2007). Toward an integrated framework of information and communication behavior: College students' information resources and media selection. In *Proceedings of the 70th annual meeting of the American society for information science and technology*, 44. Available from: http://www.asis.org/Conferences/AM07/proceedings/papers/34/34_paper.html.
- Salwen, M. B. (1987). Credibility of newspaper opinion polls: Source, source intent and precision. *Journalism Quarterly*, 64, 813-819.
- Self, C. S. (1996). Credibility. In M. Salwen & D. Stacks (Eds.), *An integrated approach to communication theory and research* (pp. 421-441). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sundar, S. S. (1999). Exploring receivers' criteria for perception of print and online news. *Journalism and Mass Communication Quarterly*, 76, 373-386.
- Sundar, S. S. (2007). Technology and credibility: Cognitive heuristics cued by modality, agency, interactivity and navigability. In M. Metzger & A. Flanagin (Eds.), *Digital media, youth, and credibility. MacArthur Foundation Series on Digital Media and Learning* (pp. 73-100). Cambridge, MA: The MIT Press.
- Toms, E. G., & Duff, W. (2002). "I spent 1 1/2 h sifting through one large box..." Diaries as information behavior of the archives user: Lessons learned. *Journal of the American Society for Information Science and Technology*, 53(14), 1232-1238.
- Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Communications of the ACM*, 42(5), 39-44.
- Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 53(2), 134-144.
- Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Westport, CT: Greenwood Press.

Digital Media and Youth: Unparalleled Opportunity and Unprecedented Responsibility

Andrew J. Flanagin and Miriam J. Metzger

University of California, Santa Barbara, Department of Communication

With the sudden explosion of digital media content and access devices in the last generation, there is now more information available to more people from more sources than at any time in human history. Pockets of limited availability by geography or status notwithstanding, people now have ready access to almost inconceivably vast information repositories that are increasingly portable, accessible, and interactive in both delivery and formation. Basic human activities have changed as a result, and new possibilities have emerged. For instance, the process by which people locate, organize, and coordinate groups of individuals with shared interests, the number and nature of information and news sources available, and the ability to solicit and share opinions and ideas across myriad topics have all undergone dramatic change as a result of interconnected digital media.

One result of this contemporary media landscape is that there exist incredible opportunities for learning, social connection, and individual entertainment and enhancement in a wide variety of forms. Indeed, recent evidence indicates that 45 percent of users in the United States say that the Internet played a crucial or important role in at least one major decision in their lives in the last two years, such as attaining additional career training, helping themselves or someone else with a major illness or medical condition, or making a major investment or financial decision.¹ Enhanced connectivity and information availability have changed not only *what* people know, but *how* they know what they know.

However, the wide-scale access and multiplicity of sources that ensure vast information availability also make assessing the credibility of information extremely complex. The origin of information, its quality, and its veracity are now in many cases less clear than ever before, resulting in an unparalleled burden on individuals to locate appropriate information and assess its meaning and relevance accurately. Doing so is highly consequential: assessing credibility *inaccurately* can have serious social, personal, educational, relational, health, and financial consequences. As a result, determining trust, believability, and information bias—key elements of credibility—become critical as individuals process the information in their lives gleaned from digital media.

Understanding credibility in this environment is also important because it is a concern that cuts across personal, social, and political domains. For instance, digital media increasingly deliver information that results (or fails to result) in an informed citizenry that, in turn, drives the pursuit of particular social agendas, the degree and nature of engagement in public discourse, and the determination of public policy.² In addition, in light of the prevalence of interactions with others that now routinely occur online, personal and social identity is often established and known primarily or exclusively through computer-mediated interactions.³

Thus, the credibility of information sources is no longer necessarily a function of sustained, face-to-face interaction, nor is it established solely through the endorsement of those whom one knows personally or directly. Nonetheless, source credibility remains a key component of persuasion, with decision-making implications ranging from consumer choices to political candidate selection.

Contemporary youth are a particularly interesting group to consider with regard to credibility today. In many ways, this generation (demarcated roughly by birth around 1980, give or take a few years) is different from any before in its relationship to information technology, and also in its relationship to information sources.⁴ Known variously as "The Millennials,"⁵ the "Net Generation,"⁶ "Digital Natives,"⁷ and "Generation Y,"⁸ members of this generation share the feature of having been immersed in an environment of digital technologies (e.g., computers and the Internet) for their entire lives. They have been described, for instance, as "digital natives in a land of digital immigrants" whose experience with digital media has led them to have "new and different expectations about how to gather, work with, translate, and share information."⁹ Compared to their elders, they are more likely to turn to digital media first when researching a topic for school or personal use, to read news on the Internet than in a printed newspaper; and to use online social network tools to meet friends and to find information. In other words, the primary sources of information in their world are often (although not exclusively) digital, which is quite different from any generation prior.

Their special relationship to digital media greatly influences the way they approach learning and research.¹⁰ As the first generation to grow up with interactive digital media, millennials are comfortable with collaborating and sharing information, and do so "in ways that allow them to act quickly and without top-down direction."¹¹ This, of course, has profound implications for credibility construction and assessment. The interactivity afforded by digital media has set up an expectation among many young people to play roles of both information source and receiver simultaneously as they critique, alter, remix, and share content in an almost conversational manner using digital media.¹² Here again, the implications of the uses of technology that are favored by young people present new and different challenges for both discerning and learning to create credible information.

Nonetheless, despite these realities, examinations of youth and digital media have often been somewhat crude, focusing for example on the popular generation gap caricature, where youth are portrayed as technologically adept compared with adults. Such considerations fail to focus on the most important and enduring by-products of heavy reliance on digital media: The impact of "growing up digital"¹³ is that more and more of the information that drives our daily lives is provided, assembled, filtered, and presented by sources that are largely unknown to us, or known to us primarily in nontraditional ways. Yet, we have only begun to explore what this means, particularly for younger users who are not only immersed in digital media now but will be for the entirety of their lives.

In terms of credibility specifically, youth are also intriguing in large part due to the tension between their technical and social immersion and skill with digital tools and their inherent limitations owing to their limited development and experience. On the one hand, those who have literally grown up in an environment saturated with digital media technologies may be highly skilled in their use of technologies to access, consume, and generate information. This view suggests that in light of their special relationship to digital tools, youth are especially well positioned to navigate the complex media environment successfully. On the other hand, youth can be viewed as inhibited, in terms of their cognitive and emotional development, life experiences, and familiarity with the media apparatus. This perspective suggests that

although youth are talented and comfortable users of technology, they may lack crucial tools and abilities that enable them to seek and consume information effectively.¹⁴

As a way of beginning to understand the complex relationship between credibility, digital media, and youth, we proceed by first defining credibility after situating it in relation to allied terms and concepts. With this background, we consider credibility in the digital media environment, and examine the various credibility assessment strategies in use today, particularly those relying on group and social engagement. From there, we consider why credibility is worthy and important as a topic of inquiry, including what is and is not new about credibility in the context of digital media. Next, we examine the major issues with regard to credibility and, in particular, what special concerns arise for youth populations. Finally, we offer some perspectives of scholars, educators, and, most important, youth regarding credibility, and with this context we examine the research, policy, and educational implications. We conclude by considering the gaps to be filled in credibility research and providing recommendations for practitioners of all kinds who are affected by youth, credibility, and digital media.

Credibility Origins, Definitions, and Current Issues

Origins of Credibility Research

Scholarly interest in credibility dates back to Aristotle's writings on rhetoric and his notions of ethos (appeal based on the character of a speaker; e.g., reputation), pathos (appeal based on emotion; e.g., fear appeals), and logos (appeal based on logic or reason; e.g., the content of a speech). According to Aristotle, all three were necessary to be trustworthy,¹⁵ although in his view credibility was mainly in the speaker and his or her ability to relate to different audiences. Ethos was thus "the communicator's ability to inspire confidence and belief in what was being said,"¹⁶ and high-ethos speakers were considered fair, trustworthy, sincere, reliable, and honest.

Source credibility was addressed in earnest during the twentieth century by psychologists interested in studying persuasion, largely as a response to propaganda efforts during the World Wars. The "Yale Group," led by social psychologist Carl Hovland, defined credibility as expertise and trustworthiness and, for the first time, drew a distinction between source credibility, message credibility, and audience credulity.¹⁷ In contrast to Aristotle's view, they suggested that credibility is a receiver-based construct and is determined by the acceptance of a speaker by the audience. The Yale group conducted numerous studies of source credibility as it pertained to persuasion and attitude change and identified the major components of what it means for a source to be perceived as credible by an audience. This work spurred a large body of research looking at both "source" and "message" credibility—characteristics of speakers and characteristics of messages or information.

The next major interest in credibility research came from professional rather than academic concerns. As television diffused widely in the 1950s, subscription rates for daily newspapers started to sag. As a result, newspaper professional organizations became interested in the perceived credibility of newspapers versus television (i.e., "media credibility"). The major finding in this domain was that the more people relied on a medium for news—television or newspapers—the more credible they believed that medium was.¹⁸

The study of credibility was resurrected in the late 1990s by the emergence of the Internet, the Web, and academic (psychology, communication, persuasion) and professional (news, e-commerce) concerns surrounding these technologies. In this domain, the Internet and

Web conflate notions of source, media, and message credibility, which formerly have been treated as rather distinct, or at least addressed largely separately.¹⁹

Defining Credibility

There exists no one, clear definition of credibility that has arisen from this heritage. Rather, the overarching view is that credibility is the *believability* of a source or message, which is made up of two primary dimensions: trustworthiness and expertise. Some secondary dimensions include source dynamism (charisma) and physical attractiveness, for example. Moreover, the two primary dimensions (trustworthiness and expertise) have both objective and subjective components. That is, trustworthiness is a receiver judgment based primarily on subjective factors. Expertise can be similarly subjectively perceived but includes relatively objective characteristics of the source or message as well (e.g., source credentials or information quality).

The study of credibility is highly interdisciplinary, and definitions are also somewhat field-specific. For example, communication and social psychology treat credibility as a perceptual variable: credibility is not an objective property of a source or a piece of information; instead, it is a subjective perception on the part of the information receiver.²⁰ Thus, this perspective emphasizes audience *perceptions* of credibility rather than the objective credibility of a source or piece of information. Information science perspectives, by contrast, view credibility as more of an objective property of information given that field's focus on defining credibility in terms of information "quality," which is how useful, good, relevant, reliable, accurate, and so forth some information is for a specific purpose. Thus, one key disciplinary distinction is that while psychology and communication tend to focus on *source credibility*, information science focuses instead on message or *information credibility*. Nonetheless, these distinctions are not perfectly clean. For example, source credibility is often a criterion for judging information credibility. Furthermore, researchers variously study the objective characteristics that make some source or piece of information *worthy* of being believed (e.g., a source's qualifications or how "good" a piece of information is compared to some accuracy metric), while others study what characteristics make sources or information *likely* to be believed by audience members (e.g., the attractiveness of a source or the congruence of some message with the receiver's own point of view).

Moreover, the notion of credibility is allied closely with several concepts, including trust, reliability, accuracy, reputation, quality, authority, and competence. Although several of these concepts include both of the core dimensions of credibility, some seem to more closely resemble the trustworthiness dimension (e.g., reputation, reliability, trust), while others seem to tilt toward the expertise dimension (e.g., quality, accuracy, authority, competence). It is beyond the scope of this discussion to elucidate the complex and intricate relationships between these concepts, but Rieh and Danielson²¹ provide a useful discussion of this issue.

Understanding Credibility in the Digital Media Environment

Many studies of credibility of Web-based information rely in some form on the traditional distinctions of source, message, and medium credibility. *Source credibility* has conventionally considered characteristics of persuasive sources (traditionally, speakers), primarily expertise and trustworthiness, but also dynamism, composure, sociability, liking for the source, and similarity to the source.²² Conceiving of Web sites as sources that engender greater or lesser credibility has shown that it is possible to translate several components of source credibility to the Web environment. Specifically, expertise may be communicated through the accuracy

and comprehensiveness of a Web site's information, its professionalism, and its sponsor's credentials.²³ Trustworthiness is associated with a Web site's integrity as demonstrated by its policy statements, use of advertising, professionalism, and firm or author reputation.²⁴ Attractiveness and dynamism, or how "entertaining" a site is, may be reflected in the site's use of colorful graphics, interesting content, or interactive features.²⁵ Finally, differences exist across Web content "sponsors": institutional Web sites are perceived as more credible than other types, for example, commercial, advocacy, and personal Web sites.²⁶

Message credibility typically examines how message or information characteristics influence perceptions of believability. Major dimensions of message credibility include message structure, content, language, and delivery.²⁷ Although little research has directly addressed message credibility online, Internet users report very similar criteria in judging online and offline information.²⁸ Aspects of message content such as accuracy, use of evidence and citations, comprehensiveness, and currency have been shown to affect perceptions of the credibility of information online.²⁹ Moreover, Fogg et al.³⁰ found that structural characteristics of online messages, such as their organization (or navigability), and message delivery elements, like the presence of typographical errors, affect credibility assessments. Looking across the literature, several message attributes appear to affect credibility perceptions of messages when they appear online.

Finally, studies of *media credibility* focus on the relative credibility or believability of various media channels through which a message is sent. Cross-media comparisons have sought to assess the credibility of digital media relative to other communication channels, with mixed results. While some studies have found that traditional mass media (e.g., newspapers) are perceived as more credible than the Internet and Web,³¹ others have found the opposite results³² or have found no differences between traditional and digital channels of information.³³ Overall, research in this area indicates that although the Web is largely considered an equally credible source of information as compared to traditional venues, it may be perceived as more credible among those who are particularly motivated to seek out specific types of information and who may rely on the Web to a large extent.³⁴

While conceptually tidy, Chaffee³⁵ argued that various dimensions of credibility overlap, and that many information consumers do not distinguish, for example, between the source of a message and the channel through which they receive the message. This type of convergence is especially pronounced in today's media environment that offers an astonishing amount of information, across various media, from a vast array of providers. Moreover, perceptions of credibility vary from person to person and between various media; digital media venues such as Web sites are themselves moving targets, constantly changing and evolving; users are also evolving, with regard to their experiences, capabilities, and the media environment in which they mature; and, there are many levels of analysis consider. For example, it makes sense to measure the credibility of the Web as a medium of communication, various forms or tools of Internet communication separately (e.g., Web sites, blogs, e-mail), entire Web sites, particular information or messages on a Web site, a site operator (e.g., nytimes.com), or individual authors of information (e.g., former *New York Times* reporter Jayson Blair). Thus, source, message, and medium credibility are overlapping concepts in many instances, and research designs that do not always enable clear distinctions among these factors complicate our current understanding of online credibility. Overall, such factors underscore the complexity of credibility in the current media environment.

Indeed, making sense of credibility today requires accounting for the various types of digital media and forms of information currently available, in order to understand how

individuals assess both information and source credibility, and how each of these influences the other. To date, however, research examining the credibility of information people obtain via digital media has primarily examined the perceived credibility of Web sites, as opposed to considering the full range of available digital information resources (e.g., e-mail, blogs, text messaging), and has tended to emphasize how individuals assess credibility in isolation, rather than considering group and social-level processes. Yet, in addition to commercial, informational, and other Web sites produced by organizations or individuals, blogs, wikis, social networking sites, and other digital media applications—linked across a wide variety of devices—constitute a significant portion of today's media environment. In fact, these tools may be especially popular among younger users.³⁶ It is crucial, therefore, to consider what new and emerging types of credibility and credibility assessment are implicated in these media tools.

To highlight this, we next propose a categorization of credibility construction, which is leveraged to draw attention to the range of relevant credibility assessment strategies available to information consumers today. This approach is intended to provide a way to organize and consider the diverse means by which information consumers understand, approach, and assess credibility in the contemporary digital media environment.

Contemporary Forms of Credibility and Credibility Assessment

A hallmark of the digital media environment is the ability of individuals to connect to one another more easily owing to reductions in the costs of communication and information sharing. Nonetheless, the majority of research on online credibility considers individuals largely as isolated appraisers of credibility, rather than as networked actors engaged with others. Group and social engagement, however, are crucial to credibility construction and assessment, and are likely increasingly important to younger user groups, which are the first to mature with a full appreciation of the potential of networked environments.

Processes of social endorsement—a fundamentally network phenomenon—have always been central to credibility. In communities where individuals and other entities are relatively well known to one another, a small number of endorsements can serve to effectively establish credibility. However, in large, relatively anonymous environments, personal connections become more tenuous. In these instances, credibility may be constructed by members of informally bounded groups of individuals who have some form of firsthand experience with the target under scrutiny, which can range from individuals, to products, to organizations or institutions, to which they lend their endorsement. In essence, endorsed credibility in the digital media environment compensates for the relative anonymity of tools like the Web with the sheer volume of users, at least some of whom have had private experiences that they make public via communication networks. The means of sharing these assessments can take many forms, resulting in several variants of credibility, most notably *conferred*, *tabulated*, *reputed*, and *emergent* credibility.

Credibility may be conferred on some information or its source when well-regarded entities, such as organizations, agencies, or associations, produce or recommend things like information repositories or service providers to information consumers. For example, libraries and teachers confer credibility on the information databases they make available to their patrons and students,³⁷ and doctors confer credibility on the sites they recommend to patients.³⁸ Similarly, organizations confer the credibility of their “preferred vendors,” and the Better Business Bureau confers credibility on those businesses that adhere to their standards of conduct. In such instances, entities establish credibility by leveraging their expertise to approve a resource.

Of course, the effectiveness of *conferred credibility* rests on the referring entity's widely recognized, positive reputation that alleviates users' skepticism. However, if users fail to recognize relationships between sources and recipients of conferrals that might compromise credibility, conferred credibility may be earned falsely. For example, the search engine Google currently provides a number of sponsored links, for which Google has been financially compensated, that appear on each page of search results. Research shows, however, that the majority of users is unaware of the sponsored status of such links and views these resources as equivalent to the remainder of search results shown.³⁹ In this case, Google has conferred its credibility to the sponsored links, if not intentionally than at least functionally, in light of users' ignorance of the preexisting sponsorship model. Similarly, hyperlinks among blogs and Web pages of similar content might indicate a presumed endorsement when instead the linkage might be based simply on shared topical interest.

Tabulated credibility relies on peer rating of some dimension of an individual, organization, transaction, opinion, or product that is subsequently tallied to provide an omnibus rating score. For example, the online auction site eBay.com relies on its members to rate others with whom they have engaged in a transaction, in order to mitigate the considerable risk involved in such financial transactions by enhancing trust, or the "perception of the degree to which an exchange partner will fulfill the transactional obligations in situations characterized by risk or uncertainty."⁴⁰ Several studies have demonstrated that tabulated group-based credibility rating systems such as the one used by eBay are consequential for users: indicators of positive reputation can result in higher bid prices, more bid activity, items that are more likely to sell, and fewer problematic transactions⁴¹ and indicators of negative reputation can result in lower bid prices or fewer bids.⁴²

Tabulations may also emerge from individuals who interact with one another indirectly. For example, amazon.com provides not only a venue for users to rate their experiences with products and retailers, but also provides means by which users can assess reviews provided by fellow Amazon users. In this manner, those providing reviews earn a reputation over time as being (non)credible information sources. Importantly, such reputational ratings demonstrate the complexity of the concept of credibility: reputation differs from credibility inasmuch as one can potentially be highly credible, but yet have a negative reputation.

Tabulation procedures can also be fairly complex. The online forum slashdot.org, for instance, provides a sophisticated group-based credibility tool to rate its users, their posts, and content provided by and to members. Especially active and valued members are given "moderator" status, whereby they can edit conversational strands and rate contributions to the forum. Moderators rotate over time and are selected from among good contributors, defined by their "karma," which is assigned based on an assessment of "good" versus "bad" posts to the forum. Although complex, this system appears to be effective among slashdot's approximately 700,000 users. From the user's perspective, the ability to aggregate ratings from known or unknown others widens the range of social input that information consumers can use to judge credibility in a way that was not possible before the appearance of networked digital media.

Reputed credibility is also a form of endorsement that is perpetuated through personal and social networks.⁴³ A good reputation is a powerful mechanism for persuasion. Individuals may not even be able to recount the origins of their reputational credibility perceptions of various sources, relying on general recollections rather than specific information or on subtle cues conditioned over experience and time. In this manner, information consumers might widely recognize and understand differences between the *New York Times* and *New York Post*, for example, in both their print and online forms. Indeed, studies show evidence that

information seekers rely on reputational cues gleaned from their social networks as a primary means of assessing credibility online.⁴⁴ Accordingly, sources strive to construct credibility by establishing a good reputation and perpetuate that reputation through social networks. In addition, there is evidence for “credibility transfer,” where “recipients use the credibility of a medium or media product as a (heuristic) indicator for the credibility of a single news story or programme.”⁴⁵ In this manner, credibility transfer can occur both between different media (e.g., from the *New York Times* printed version to NYTimes.com) and within a medium (e.g., from a credible Web site such as CNN.com to a story residing on that site).

Finally, *emergent credibility* also arises from group and social engagement. Several online venues including wikis, social networking sites, and many other applications now provide vast information repositories created by largely self-coordinating individuals, rather than by a central organization. The result is that credibility can sometimes be an emergent phenomenon that arises from a pool of resources, achieved through a system of open access to all. Emergent credibility is at the core of Lankes’s⁴⁶ “reliability approach” to credibility and Eysenbach’s⁴⁷ discussion of “apomediaries” as the new arbiters of credibility, particularly among youth who are at ease with social uses and applications of digital media.⁴⁸ Wikipedia.com, the vast online encyclopedia with over one million entries provided entirely by individual users, is a prime example of emergent credibility. Indeed, a direct comparison of the accuracy of science entries in Wikipedia and Encyclopedia Britannica revealed very few differences.⁴⁹ Ultimately, credibility through emergence draws attention to crucial elements of credibility assessment: in such environments, users would be wise to consider issues of bias, source identity, and perspective as they navigate resources provided collectively by multiple interested parties.

Concerns about Credibility and Digital Media

Concerns about the credibility of sources and information certainly pre-date the advent of digital media. The need to teach young people to critically appraise information has long been a part of educational efforts, under various monikers such as literacy training and critical thinking.⁵⁰ In many ways, the core skills and issues in this domain are the same today as they were before the recent rise in digital technologies. As argued elsewhere, digital media have not so much changed what skills are needed to evaluate the credibility of information as they have changed the need for people to know how and when to exercise those skills.⁵¹

Digital media do, however, present new challenges for information consumers, and have in many ways shifted the burden of information evaluation from professional gatekeepers to individual information consumers. Accordingly, several scholars have addressed the question of what is new about digital media that makes the need for effective critical evaluation more pressing today.⁵² This discussion focuses on why digital media present special problems with regard to credibility and credibility assessment, including the quantity and access of information afforded by digital media technologies, the lack of gatekeepers and quality control standards, source and context ambiguity, convergence of information and media channels, disintermediation, and shifting norms and expectations for information retrieval and processing. These functions are examined next.

Special Circumstances of Digital Media and Credibility

Today, few question the notion that digital, networked media have profoundly changed the information landscape, as well as the means of social interaction. Perhaps the greatest change

is that digital media have provided access to an unprecedented amount of information available for public consumption. Until recently, the enormous cost and complexity involved in producing and disseminating information limited the number of information providers, who generally had substantial financial investment in the media apparatus.⁵³ Network and digitization technologies, however, have lowered the cost of information production and dissemination, thus increasing the sheer amount of information available. Indeed, the adage “On the Internet, anyone can be an author” is largely true, barring, of course, a few well-known caveats about cost and access to technology.

The combination of the vast quantity of and accessibility to digitally stored and transmitted information has prompted concerns about its credibility because, as Rieh and Danielson⁵⁴ argue, this combination creates greater uncertainty regarding both who is responsible for information and, consequently, whether it can be believed. Two important and related issues are the nature of gatekeeping in the digital media environment and the level of ambiguity surrounding both the source and context of information.

Several scholars have pointed out that information posted on the Web may not be subject to filtering through professional gatekeepers and, as a result, digital information may be more prone to being poorly organized, out of date, incomplete, or inaccurate.⁵⁵ Others have noted that digital media sometimes lack traditional authority indicators such as author identity or established reputation.⁵⁶

Indeed, source information is crucial to credibility because it is the primary basis on which credibility judgments are thought to rest. At the same time, however, “source” has become muddled as media shifted from analog to digital forms.⁵⁷ There are several ways in which the source of information is problematic in the digital media environment. In some cases, source information is unavailable, masked, or entirely missing from a Web site, chat group, blog, wiki, and so on. In other cases, source information is provided, yet hard to interpret, such as when information is coproduced; re-purposed from one site, channel, or application to another; or when information aggregators display information from multiple sources in a centralized location that may itself be perceived as the source. These technological features create a kind of “context deficit” for digital information.⁵⁸ Moreover, the hyperlinked structure of the Web contributes to this deficit by making it psychologically challenging for users to follow and evaluate various sources as they move from site to site. Research by Eysenbach and Kohler,⁵⁹ for example, showed that source and message information become confused or disassociated in users’ minds almost immediately after performing searches for medical information online.

Concerns about credibility within the digital media environment also stem from the fact that there are few standards for quality control and evaluation. There are no universal standards for posting information online, and digital information may be easily altered, plagiarized, misrepresented, or created anonymously under false pretenses. The malleability and dynamic nature of digital information exacerbate potential problems of information reliability, given that the alteration of digital information is difficult—if not impossible—to detect. In addition, the global nature of the Web makes it challenging to enact standards for quality control in the form of government regulation.⁶⁰ Finally, there is greater ambiguity about how to evaluate digital information owing simply to the relative newness of these channels of information that, in turn, makes the lack of standards for information presentation and evaluation more significant in comparison to traditional media.⁶¹

Another reason that the credibility of digital information may be suspect relative to more traditional media is due to channel convergence and conflation of content types afforded

by digital technologies. Some have suggested that visual and other types of distinctions that were once clear between, for example, information and commercial content are not so easily distinguished in the digital environment.⁶² Examples include sponsored and unsponsored links on search engine result pages and ads embedded in Web page content. Indeed, Burbules⁶³ has suggested that because information is presented in a similar format on Web sites, a psychological "leveling effect" is created that puts all information on the same level of accessibility and, thus, all sources on the same level of credibility.

In the extreme, "spam" messages (unsolicited or inappropriate messages often used for mass commercial advertising) create instances where individuals lack knowledge of the sender, although peripheral cues can serve to inform credibility evaluations. Less obvious, however, are other hybrid e-mail forms. "Phishing" techniques (e-mail messages from presumably known sources that are designed to entice individuals to visit fraudulent Web sites) are designed to appear credible in order to elicit individuals to provide personal data such as bank account information, and have been shown to be very effective, in spite of widespread knowledge of such techniques.⁶⁴ In addition, as "viral" e-mails are forwarded from one person to another, the content of the message can sometimes be so thoroughly divorced from its source as to make evaluation nearly impossible. Interestingly, spam, phishing, and viral e-mails can be viewed as attempts to construct credibility by capitalizing on users' perceptions of the credibility of the e-mail medium, which is derived in large part by its highly personal nature, as most e-mail communication takes place between parties known to one another.

Finally, digital media prompt increased concern about credibility by elevating the negative consequences of misinformation for consumers. As part of their efforts to economize, organizations are migrating more and more critical information to the Web, or making such information accessible exclusively via digital means.⁶⁵ Digital media have thus enhanced both capabilities and expectations for people to be more self-serving and self-informing. Individuals are now encouraged or expected to do everything from choose between medical treatment options, decide on retirement benefits and investment options, book airline reservations, and select and register for college courses entirely on their own, using information provided via digital media rather than interacting with live agents, experts, or customer service representatives. This trend toward "disintermediation" enabled by digital media raises the stakes for consumers to be able to locate and discern credible information online.⁶⁶

Overall, while it is true that these characteristics and realities of digital media may not have changed the basic skills needed for credibility assessment, they certainly have changed the *need* to assess credibility, the *frequency* with which to do so, and the *strategies* that may be useful and available to assess information and its source. As Burbules⁶⁷ notes, "conventional methods for assessing credibility may not be feasible on the Web because of its speed, complex features and link structure, and lack of referencing and organizational conventions."⁶⁸

Youth, Credibility, and Digital Media

Digital media present special credibility issues for youth. To date, however, the vast majority of research on credibility has focused on news, health/medical, and commercial information, which are often assumed to be "adult" topics, or at least topics that are of more interest to adults. Consequently, extremely little research has focused on credibility and youth, in spite of the fact that youth are avid information seekers across many domains of information, whether for class assignments or personal use.⁶⁹ It may be the case, however, that youth seek different *types* of information using digital media than do adults. For example, Eysenbach⁷⁰

points out that while adults often seek medical information about treatments or disease, youth are more likely to seek information on health (e.g., fitness, dieting) or sexuality.

The credibility of information obtained via digital media is important for youth, however, not only because they are active information seekers but also because there are some key differences between youth and adult information seekers in general, and specifically in their information-seeking goals. Compared with adults, for example, youth tend to be relatively heavily immersed in digital media by virtue of growing up in an environment saturated with these tools.⁷¹ There is also evidence that youth access an equal or greater proportion of information via digital media than do adults, suggesting that they may find using these tools to locate information more natural compared with adults.⁷²

Indeed, librarians and educators around the country have commented on a significant decline of (physical) library patronage since the appearance of the Internet and searchable databases.⁷³ Martell documents these trends with statistics across various university libraries, showing corresponding declines in physical library usage with increases in remote, virtual usage.⁷⁴ Beyond locating information for academic purposes, youth also rely heavily on digital media for other informational purposes on a daily basis, using a wide array of digital media technologies such as social networking Web sites, chat groups, interactive games, cellular telephones, e-mail, and text messaging to do so.⁷⁵

Such heavy reliance on digital media could also be a consequence of the fact that youth may perceive greater social pressures to use digital media in ways consistent with their peers than do adults. Recent studies demonstrate that social and group-based applications of digital media, such as social networking sites, are extremely popular among young people,⁷⁶ and youth may feel greater desire and opportunity to expand their social connections and locate opinion leaders via these media, thus perhaps taking better advantage of opportunities for "emergent credibility" construction and assessment, as discussed earlier. At the same time, however, youth may also perceive less risk of disclosing personal information while using these applications, because they might think of themselves as having less financial and identity consequences at risk than adults. This presents complex credibility issues with regard to discerning the trustworthiness of some person with whom one is interacting via these media that are not unique to youth, but could be more pressing and widespread for this population.

Children's relative lack of life experience as compared to adults may also put them at greater risk for falsely accepting a source's self-asserted credibility, since such assessments are based on accumulated personal experience, knowledge, reputation, and examination of competing resources. As a group, youth have fewer life experiences to which they might compare information than do most adults. In addition, youth may not have the same level of experience with or knowledge about media institutions, which might make it difficult for them to understand differences in editorial standards across various media channels and outlets compared with adults who grew up in a world with fewer channels and less media convergence. As a consequence, some youth may not have the same level of skepticism toward digital media as adults do, because these media are not seen as "new" to younger users who cannot remember a time without them.

Specific instantiations of digital media technology may also pose particular credibility challenges for children, who possess less knowledge, experience, and skepticism than adults. E-mail serves as a good example. In the case of spam, younger children in particular might blindly transfer their perceptions of e-mail as a personal, credible medium and therefore not fully appreciate the sender's commercial intent. Similarly, children may not have sufficient

knowledge to be appropriately skeptical of phishing techniques or viral e-mail messages that make it difficult to determine the true message source and purpose. Also, differences in life experience may affect youth's ability to appropriately understand newer forms of credibility construction and assessment discussed earlier. Specifically, both "reputed" and "conferred" credibility rest on users' familiarity with the reputation of a particular source or conferring agent, a familiarity that many children may not possess. Furthermore, it may be difficult for children to grasp the often complex relationships between sources and recipients of credibility conferrals, such as the meaning of "sponsored links" in a Google search result page.

Beyond experiential differences, children differ from adults in their cognitive development. As described by Eastin,⁷⁷ there are clear differences between children and adults, and between younger and older children, in cognitive skill acquisition that may have important credibility implications. Youth, particularly younger children, may be more susceptible to digital misinformation and less able to discern credible from *noncredible* sources and information than are adults who are more cognitively advanced. Evidence shows that young children have a hard time distinguishing commercial from noncommercial information within the broadcast medium,⁷⁸ a task that can only be more difficult given trends toward channel convergence and conflation of types of information in the digital media environment discussed earlier. Other credibility "aids" that may be effective for adults may be ineffective for children who have not reached cognitive maturity. For example, a popular tool to help people negotiate unfamiliar territory online is ratings and recommender systems. The intricacies of such systems might be too difficult for younger children to grasp since these systems can actually be quite complex. To take full advantage of ratings systems, for example, users must weigh knowledge about the quality and quantity of raters, the context in which ratings are proffered, and the specific tabulation system in place. This suggests that the advantages offered by "tabulated credibility" may be lost on younger digital media users.

Differences in reliance, motivation for and patterns of use, experience, and development notwithstanding, both children and adults face serious consequences of receiving unreliable information from digital media sources. These consequences may, however, be somewhat different for the two user populations. For adults, the financial or health ramifications of making decisions based on bad or outdated information may be of primary concern, whereas for children the consequences of misinformation may be more apparent in other realms, such as in learning or physical safety. Learning and, by extension, grades are crucial concerns for youth, given their importance as determinants of opportunities later in life. Ongoing news reports of kidnappings and assault highlight the importance of children's ability to assess whether those to whom they disclose personal information via digital media really are who they claim to be. In each case, the cost of assessing credibility inaccurately is highly consequential for youth.

Although differences in experience and cognitive development suggest that there are substantial challenges for youth to assess the credibility of digital sources and information, it is simplistic to conclude that youth are inherently disadvantaged compared with adults when it comes to credibility assessment. These characteristics of youth audiences present both challenges and opportunities. On the one hand, youth may be a particularly vulnerable audience because of their special characteristics. On the other hand, forms of credibility assessment that rely on information to be spread efficiently through social networks (e.g., emergent credibility) highlight some intriguing advantages for youth populations, who are often extremely interconnected compared to adults. In such instances, younger users may actually be better equipped than adults to transmit information pertaining to an entity's credibility quickly and efficiently via their social networks.⁷⁹

What is safe to conclude is that youth's high degree of immersion, coupled with areas of naïveté, differences in cognitive development, and usage of digital media elevate the importance of understanding credibility within this user population. A primary purpose of this volume is, thus, a call to arms to researchers, educators, policy makers, and others concerned with these issues to understand how youth think about credibility in the digital media environment and to devise a plan to assist youth in finding and evaluating the information that they need.

Perspectives on Credibility from Scholars, Educators, and Youth

Scholars, educators, and youth bring different perspectives about digital media and youth, each informed by different concerns and foci. Collectively, these views combine to shed light on a number of relevant issues; yet, it is worthwhile to examine each of these voices singly as well.⁸⁰

Scholars

Many intellectuals point out that digital media have changed not only people's access to information but also the ways in which we appraise information. Traditional notions of credibility as coming from a centralized authority (e.g., a teacher, expert, or author) and individualized appraisal processes are challenged by digital technologies.⁸¹ Electronic networks make it easier to rely on the collective to assess information. Credibility assessments constructed through collective or community efforts (e.g., wikis, text messaging via cell phones, or social networking applications) emerge as a major theme in recent discussions, and phrases like "distributed" and "decentralized" credibility, the "democratization of information," and "collectively versus institutionally derived credibility" are common. At core is the belief that digital media allow for the uncoupling of credibility and authority in a way never before possible. Digital media thus call into question our conceptions of authority as centralized, impenetrable, and singularly accurate and move information consumers from a model of single authority based on hierarchy to a model of multiple authorities based on networks of peers.

For some, this change is scary while for others it is liberating. Indeed, there are two opposing reactions to the dangers posed by noncredible digital information. One reaction calls for "protectionist measures" involving censorship and restricted access to online information (e.g., filters or restrictive policies), which necessarily rely on systems of hierarchy and authority (e.g., portals and professional vetting of sites). For others, this shift is positive and is seen as the "solution" to the credibility "problem." This view advocates allowing *greater* openness and freedom of information and is evocative of John Milton's First Amendment principle that the best way to counter bad speech is to allow more speech. In this view, all perspectives should be allowed into the marketplace of ideas and, through that process, "bad" (i.e., noncredible) ideas will ultimately be discredited and discarded.⁸²

It is interesting that both views are evident in discussions of digital media and credibility. Educational efforts have more often taken the protectionist approach, and have opted for forms of censorship to shield students from potentially "bad" information online. Others argue that censorship and filtering hamper students' ability to learn to think critically about the information they receive via digital media,⁸³ and that collaborative filtering processes made possible by digital networked technologies will solve many of the credibility problems raised by digital technologies, as everything will be subject to unprecedented levels of peer review via digital networks.

Educators

The notion that digital media are challenging our fundamental ideas about learning and education is prevalent among teachers and librarians. Specifically, many educators argue that digital, networked media provide new opportunities for students to learn from others in a much less hierarchical and institutional way than ever before. Some go so far as to suggest that schools and teachers are no longer needed since digital media enable students to access and learn from the collective experience and intelligence of peers or communities. Of course, this type of self-directed learning carries with it a greater burden for students to critically evaluate information from these sources. The urgency with which educators speak of digital literacy is exacerbated by their feeling that digital media have changed youths' expectations about information. They say that today, young people expect information to be available at any time, instantly. Many educators feel this places greater burden on students to evaluate the information they get online, presumably because they are processing more information more quickly, and because source identity may be harder to discern online.

Some educators feel strongly that learning to assess the credibility of digital media and information requires students to participate in the online communities from which they seek information. Others feel that a critical component of digital literacy requires kids to learn credibility assessment through digital media production, thereby teaching kids to be "architects of credibility." In other words, many feel that youth need to be taught how to be "prosumers" (i.e., both producers and consumers) of digital media as a starting point for learning credibility assessment.

Another, related theme is that educators need to find ways to support what kids are naturally doing with digital media and to leverage that into opportunities to teach critical thinking. This likely involves developing new models for teaching and learning, but a problem is that most schools cast a wary eye on incorporating new forms of digital media into the curriculum. One example is that most of youths' favorite applications and uses of digital media, including social networking applications, e-mail, cell phones, and e-commerce, are banned in schools. In fact, most school policy toward digital media, as well as most media literacy curricula, is based on the assumption that children are in need of protection from vast amounts of misinformation online and other dangers lurking on the Web.⁸⁴

Youth

In contrast to the other stakeholders, youth themselves do not seem to be terribly concerned about credibility. What few empirical studies exist find that credibility is not a primary concern to young people when using digital media,⁸⁵ although two dimensions of minor apprehension have emerged. First, youth report that they are concerned about the reliability/accuracy of information that they find and use in their schoolwork (i.e., information or message credibility). Second, youth say that they are concerned about trusting people they may meet in the virtual world. Neither of these concerns is surprising, given that academic and social motivations prompt a good proportion of young people's use of digital media. As for solutions to the credibility "problem," many youth indicate that it is up to each individual to use caution and take the necessary steps of cross-validation and background research to verify digital sources and information. Whether they will rise to that challenge in every situation is both unknown and unlikely.⁸⁶

Perhaps the most consistent theme across all these stakeholders is that digital technologies complicate traditional notions of hierarchy and authority structures. The digital media environment offers unprecedented flexibility, allowing kids to have more authority than

adults in some realms on some topics, nonexperts to be more credible than experts in certain circumstances, and unknown individuals to sometimes be more believable than the national news media. Many believe that digital media are shattering traditional models of authority and, as a result, credibility is being turned on its head. This upheaval suggests both problems and opportunities that require going beyond polemics in thinking about these issues.

Conclusions and Directions for Future Inquiry

Theoretical development provides researchers and practitioners a better understanding of the processes of credibility assessment that, in turn, aids them in devising strategies to help people become better information consumers. Although research on credibility and digital media to date has often neglected theory construction,⁸⁷ important theoretical applications and developments are currently emerging, many of which are represented in this volume. In particular, researchers are beginning to recognize the role of context and motivation in information searches and to incorporate this into models of credibility assessment.⁸⁸ In this regard, the application of dual processing theories, such as the Elaboration Likelihood Model or the Heuristic-Systematic Processing Model,⁸⁹ is particularly promising since these theories can help explain differences in credibility assessment processes across a variety of information-seeking contexts and situations.

With regard to youth specifically, theories of human development offer insight into the special problems that young people of varying ages may face when assessing the credibility of sources or information in the contemporary media environment. Eastin,⁹⁰ for instance, notes the cognitive developmental limitations that potentially inhibit youth of various ages from discerning information source, intent, and therefore credibility appropriately. In this area, the interdisciplinary nature of credibility research offers particularly exciting opportunities for innovative theoretical developments.

Credibility assessment research is also starting to profit from information-processing theories, such as the Limited Capacity Model and others, that help to understand how people evaluate and make decisions based on information they obtain from digital media.⁹¹ Rieh and Hillgoss,⁹² for example, propose that individuals make predictive evaluations of the usefulness of information sources and information-seeking strategies based on their own experiences and refine these assessments over time. In this manner, people learn to reuse or avoid information sources based on their verification of them. Similar to Sundar's⁹³ observation that individuals heavily invoke cognitive heuristics in their credibility assessments, rather than more laborious information-processing activities, such strategies result in increased predictability and reduced cognitive effort. In contrast to cognitive heuristics, however, predictive/evaluative strategies rely primarily on explicit and articulated social outcomes, as opposed to largely implicit evaluations of technical features and inferred social cues formed over time and experience. To gain the greatest benefit from information-processing models, future work should combine both of these perspectives.

Eysenbach similarly argues that people's evaluative strategies evolve with experience.⁹⁴ He proposes an iterative view of technology usage for credibility assessment whereby people shift from relying on information "intermediaries" (sources that stand between information consumers and pertinent information or services, such as librarians or doctors) to using "apomediaries" (sources that mediate without standing in between consumers and pertinent information or services, such as collaborative filtering via peers or technical tools) as they gain knowledge and self-efficacy. To have the greatest traction, the specific circumstances

under which apomediation might occur will need to be articulated and validated. Moreover, as suggested by this view, the role of various group processes needs to be more fully expressed in credibility research as well. As it stands, most research focuses on the cognitive processes of individuals acting alone (although with tacit awareness of others) to assess the credibility of information they get via digital media. To better reflect how people are increasingly using digital media, the explicitly social, networked circumstances of credibility assessment need to be addressed.

Nonetheless, experience with digital tools appears to be a necessary, although not sufficient, condition to take full and appropriate advantage of the many sources of information accessed via digital media. For instance, the nearly automatic heuristic judgments learned through interaction with digital media are honed through experience;⁹⁵ accumulated experience can result in reliable information search results across various media;⁹⁶ and Internet experience has been shown by some to be positively related to assessments of the credibility of Web-based information and to verification behaviors.⁹⁷ Experience, however, should not be confounded with expertise, nor should experience with Web technologies be equated to life experiences. Each is consequential in its own way for youth determining the credibility of online sources and information. For instance, technological experience alone does not enable individuals to discern credible from noncredible information (although it can help),⁹⁸ nor can those without sufficient life experiences make sense of certain information dilemmas they may confront online. As Eastin⁹⁹ argues, youth are faced with rather complex cognitive tasks online, and youth at different developmental stages are equipped to different degrees to resolve these dilemmas effectively. Nonetheless, youth may enjoy certain advantages due to their relative immersion in digital media, which might not yet be readily apparent. This, of course, suggests a wealth of possibilities with regard to our understanding of credibility assessment that can only be understood as data are collected over long periods of time. Indeed, as noted earlier, we have only begun to explore what "growing up digital" means for media users who will be immersed in digital media for their entire lives.

From a practical standpoint, understanding the processes of credibility assessment among youth is critical. Credibility is a cornerstone of people's interactions, personal representation, academic and professional performance, and democratic expression and choice. As more delivery mechanisms become available, more information content migrates online, and more of the world's population is interconnected, it is crucial that individuals understand the implications of their media environment and learn to assess credibility in ways appropriate to their situational demands. This is especially significant for youth, who are uniquely and simultaneously advantaged and disadvantaged by their relation to contemporary media technologies.

Teaching youth about credibility assessment, then, must both make use of their existing knowledge of contemporary technologies and address their outstanding deficiencies. As Rainie¹⁰⁰ argues, research and learning among youth are increasingly self-directed, interactive with networks of peers, and reliant on group outreach and knowledge. Consequently, opportunities that use the lessons inherent in the tools that already engage youth offer strategies for teaching critical distinctions that may result in appropriate digital media literacy.¹⁰¹ For example, showing students existing Web sites whose focus is to reveal published misinformation, prompting youth to examine Wikipedia's "talk pages" where collaborators discuss contested encyclopedia entries, encouraging youth to become information providers themselves, and using direct comparisons of competing news accounts all present means by which digital tools and resources can be leveraged to illustrate credibility assessment concerns and

strategies.¹⁰² These methods offer opportunities for demonstrating core credibility issues by invoking real-time and real-world instances of contested content and behaviors, via the media technologies and tools that youth naturally enjoy using. In this manner, even “managed” environments like schools can take advantage of so-called autonomous, informal learning environments and opportunities found online, which may themselves be more credible to youth.¹⁰³

Nonetheless, in the end such strategies must overcome the “structural” challenges that are built into the political and cultural schooling environment.¹⁰⁴ In addition, issues of credibility and new media are subject to a range of supportive to restrictive sociopolitical interventions¹⁰⁵ that can alter their form and effectiveness. Ultimately, the relation among youth, digital media, and credibility today is sufficiently complex to resist simple explanations. This volume represents a first step toward mapping that complexity and providing a basis for future work that seeks to find explanations that will ultimately help scholars, educators, policy makers, and youth take advantage of the new opportunities for empowerment and learning offered by digital networked media.

Notes

1. John Horrigan and Lee Rainie, When Facing a Tough Decision, 60 Million Americans Now Seek the Internet's Help: The Internet's Growing Role in Life's Major Moments, 2006, <http://pewresearch.org/obdeck/?ObDeckID=19> (retrieved October 13, 2006).
2. See Lance Bennett, ed., *Civic Life Online: Learning How Digital Media Can Engage Youth*, The MacArthur Initiative on Digital Media and Learning, 2007.
3. See David Buckingham, ed., *Youth, Identity, and Digital Media*, The MacArthur Initiative on Digital Media and Learning, 2007.
4. Matthew Grimm, ... 'Bout your G-g-generation (Generation Y), *American Demographics* 25, no. 7 (2003): 38–41; Lee Rainie, Life Online: Teens and Technology and the World to Come (keynote address to the annual conference of the Public Library Association, Boston, MA, March 23, 2006), <http://www.pewinternet.org/ppt/Teens%20and%20technology.pdf> (retrieved November 7, 2006).
5. Neil Howe and William Strauss, *Millennials Rising: The Next Great Generation* (New York: Vintage Books, 2000).
6. Don Tapscott, *Growing Up Digital: The Rise of the Net Generation* (New York: McGraw-Hill, 1997).
7. Marc Prensky, Digital Natives, Digital Immigrants, *On the Horizon* 9, no. 5 (September/October, 2001): 1–6.
8. Grimm, ... 'Bout your G-g-generation (Generation Y).
9. Rainie, Life Online; see also Prensky, Digital Natives, Digital Immigrants, 1–6.
10. Rainie, Life Online.
11. Ibid., 7.
12. Ibid.
13. Tapscott, *Growing Up Digital*.
14. Matthew Eastin, Toward a Cognitive Developmental Approach to Youth Perceptions of Credibility, this volume.

15. Gary C. Woodward and Robert E. Denton Jr., *Persuasion & Influence in American Life*, 4th ed. (Prospect Heights, IL: Waveland, 2000).
16. Charles C. Self, Credibility, in *An Integrated Approach to Communication Theory and Research*, eds. Michael B. Salwen and Don W. Stacks (Mahwah, NJ: Erlbaum, 1996), 421–41.
17. Carl I. Hovland, Irving L. Janis, and Harold H. Kelley, *Communication and Persuasion* (New Haven, CT: Yale University Press, 1953).
18. Richard F. Carter and Bradley S. Greenberg, Newspapers or Television: Which Do You Believe? *Journalism Quarterly* 42 (1965): 29–34; Bruce H. Westley and Werner J. Severin, Some Correlates of Media Credibility, *Journalism Quarterly* 41 (1964): 325–35.
19. See Miriam J. Metzger, Andrew J. Flanagin, Karen Eyal, Daisy R. Lemus, and Robert M. McCann, Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment, *Communication Yearbook* 27 (2003): 293–335, for more detail.
20. B. J. Fogg and Hsiang Tseng, The Elements of Computer Credibility, *Proceedings of CHI '99, Human Factors in Computing Systems* (1999): 80–87; Albert C. Gunther, Biased Press or Biased Public? Attitudes toward Media Coverage of Social Groups, *Public Opinion Quarterly* 56 (1992): 147–67.
21. Soo Young Rieh and David R. Danielson, Credibility: A Multidisciplinary Framework, in *Annual Review of Information Science and Technology* 41, ed. Blaise Cronin (Medford, NJ: Information Today, 2007), 307–64.
22. David Berlo, James Lemert, and Robert Mertz, Dimensions for Evaluating the Acceptability of Message Sources, *Public Opinion Quarterly* 33 (1969): 563–675; Robert H. Gass and John S. Seiter, *Persuasion, Social Influence, and Compliance Gaining* (Boston: Allyn & Bacon, 1999); Hovland, Janis, and Kelley, *Communication and Persuasion*; William E. Jurma, Evaluations of Credibility of the Source of a Message, *Psychological Reports* 49 (1981): 778; James C. McCroskey, Scales for the Measurement of Ethos, *Speech Monographs* 33 (1966): 65–72; Robert M. Perloff, *The Dynamics of Persuasion* (Hillsdale, NJ: Erlbaum, 1993).
23. B. J. Fogg, C. Soohoo, David R. Danielson, Leslie Marable, Julianne Stanford, and Ellen R. Trauber, How Do Users Evaluate the Credibility of Web Sites? A Study with Over 2,500 Participants (paper presented at the Designing for User Experiences, San Francisco, CA., 2003).
24. Janet E. Alexander and Marsha A. Tate, *Web Wisdom: How to Evaluate and Create Information Quality on the Web* (Hillsdale, NJ: Erlbaum, 1999); Mary J. Culnan and Pamela K. Armstrong, Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation, *Organization Science* 10, no. 1 (1999): 104–15; Fogg et al., *How Do Users Evaluate the Credibility of Web Sites?*
25. Qimei Chen and William D. Wells, Attitude toward the Site, *Journal of Advertising Research* 39, no. 5 (1999): 27–37; Eastin, Toward a Cognitive Developmental Approach.
26. Andrew J. Flanagin and Miriam J. Metzger, The Perceived Credibility of Web Site Information as Influenced by the Sex of the Source, *Computers in Human Behavior* 19 (2003): 683–701; Andrew J. Flanagin and Miriam J. Metzger, The Role of Site Features, User Attributes, and Information Verification Behaviors on the Perceived Credibility of Web-based Information, *New Media and Society* 9, no. 2 (2007): 319–42.
27. E.g., Gass and Seiter, *Persuasion, Social Influence, and Compliance Gaining*; M. A. Hamilton, Message Variables That Mediate and Moderate the Effect of Equivocal Language on Source Credibility, *Journal of Language and Social Psychology* 17 (1998): 109–43; James C. McCroskey, A Summary of Experimental Research on the Effects of Evidence in Persuasive Communication, *The Quarterly Journal of Speech* 55 (1969): 169–76; James C. McCroskey and R. Samuel Mehrley, The Effects of Disorganization and Nonfluency on Attitude Change and Source Credibility, *Speech Monographs* 36 (1969): 13–21; Gerald R. Miller and

Murray A. Hewgill, The Effect of Variations in Nonfluency on Audience Ratings of Source Credibility, *Quarterly Journal of Speech* 50 (1964): 36–44; Daniel J. O'Keefe, *Persuasion: Theory and Research* (Newbury Park, CA: Sage, 1990).

28. See Metzger et al., Credibility for the 21st Century, for a review.

29. Soo Young Rieh and Nicholas J. Belkin, Understanding Judgment of Information Quality and Cognitive Authority in the WWW, in *Proceedings of the 61st Annual Meeting of the American Society for Information Science* 35, ed. Cecilia M. Preston (Medford, NJ: Information Today, 1998), 279–89; S. Shyam Sundar, Effect of Source Attribution on Perception of Online News Stories, *Journalism and Mass Communication Quarterly* 75, no. 1 (1998): 55–68.

30. Fogg et al., *How Do Users Evaluate the Credibility of Web Sites?*

31. Andrew J. Flanagin and Miriam J. Metzger, Perceptions of Internet Information Credibility, *Journalism and Mass Communication Quarterly* 77, no. 3 (2000): 515–40; Spiro Kioussis, Public Trust or Mistrust? Perceptions of Media Credibility in the Information Age, *Mass Communication & Society* 4, no. 4 (2001): 381–403.

32. Thomas J. Johnson and Barbara K. Kaye, Cruising Is Believing? Comparing Internet and Traditional Sources on Media Credibility Measures, *Journalism and Mass Communication Quarterly* 75, no. 2 (1998): 325–40; Thomas J. Johnson and Barbara K. Kaye, Wag the Blog: How Reliance on Traditional Media and the Internet Influence Credibility Perceptions of Weblogs among Blog Users, *Journalism and Mass Communication Quarterly* 81, no. 3 (2004): 622–42.

33. Online News Association, Digital Journalism Credibility Survey, 2001, <http://www.journalists.org/Programs/ResearchText.htm> (retrieved June 25, 2001).

34. Thomas J. Johnson and Barbara K. Kaye, Using Is Believing: The Influence of Reliance on the Credibility of Online Political Information among Politically Interested Internet Users, *Journalism and Mass Communication Quarterly* 77, no. 4 (2000): 865–79; Thomas J. Johnson and Barbara K. Kaye, Webelievability: A Path Model Examining How Convenience and Reliance Predict Online Credibility, *Journalism and Mass Communication Quarterly* 79, no. 3 (2002): 619–42.

35. Steven H. Chaffee, Mass Media and Interpersonal Channels: Competitive, Convergent, or Complementary? in *Inter/Media: Interpersonal Communication in a Media World*, ed. Gary Gumpert and Robert Cathcart (New York: Oxford University Press, 1982), 57–77.

36. Amanda Lenhart, Mary Madden, and Paul Hitlin, *Teens and Technology: Youth Are Leading the Transition to a Fully Wired and Mobile Nation*, Pew Internet & American Life Report (July 2005), http://www.pewinternet.org/report_display.asp?r=162 (retrieved April 17, 2006).

37. See Soo Young Rieh and Brian Hilligoss, College Students' Credibility Judgments in the Information Seeking Process, this volume.

38. See Gunther Eysenbach, Credibility of Health Information and Digital Media: New Perspectives and Implications for Youth, this volume.

39. Deborah Fallows, *Search Engine Users: Internet Searchers Are Confident, Satisfied and Trusting—But They Are Also Unaware and Naïve*, Pew Internet & American Life report, January 2005, http://www.pewinternet.org/PPF/r/146/report_display.asp (retrieved April 17, 2006).

40. Brian P. Bailey, Laura J. Gurak, and Joseph A. Konstan, Trust in Cyberspace, in *Human Factors and Web Development*, ed. Julie Ratner (London: Erlbaum, 2003), 311–21.

41. Patrick Bajari and Ali Hortacsu, The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions, *The Rand Journal of Economics* 34, no. 2 (2002): 329–55; Andrew

- J. Flanagin, Commercial Markets as Communication Markets: Uncertainty Reduction through Mediated Information Exchange in Online Auctions, *New Media & Society* 9, no. 3 (2007): 401–23; Daniel Houser and John Wooders, Reputation in Auctions: Theory, and Evidence from eBay, University of Arizona, 2000, http://info-center.ccit.arizona.edu/~econ/working-papers/Internet_Auctions.pdf (retrieved October 1, 2003); Cynthia G. McDonald and V. Carlos Slawson, Reputation in an Internet Auction Market, *Economic Inquiry* 40 (2002): 633–50; Mikhail I. Melnik and James Alm, Does a Seller's ECommerce Reputation Matter? Evidence From eBay Auctions, *The Journal of Industrial Economics* 50, no. 3 (2002): 337–50; Paul Resnick and Richard Zeckhauser, Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System, in *Advances in Applied Microeconomics*, vol. 11, *The Economics of the Internet and E-Commerce*, ed. Michael R. Baye (Amsterdam: Elsevier Science, 2002); Stephen S. Standifird, Reputation and E-Commerce: eBay Auctions and the Asymmetrical Impact of Positive and Negative Ratings, *Journal of Management* 27 (2001): 279–95.
42. Bajari and Hortacsu, The Winner's Curse, Reserve Prices, and Endogenous Entry; Ulrich Brinkman and Mathias Siefert, Face-to-Interface—The Establishment of Trust in the Internet: The Case of E-Auctions, *Journal of Sociology* 30 (2001): 23–47; David Lucking-Reiley, Doug Bryan, Naghi Prasad, and Daniel Reeves, Pennies from eBay: The Determinants of Price in Online Auctions, Vanderbilt University working paper, 2000, <http://eller.arizona.edu/%7Ereiley/papers/PenniesFromEBay.pdf> (retrieved August 5, 2006); Standifird, Reputation and E-Commerce.
43. See Shawn Tseng and B. J. Fogg, Credibility and Computing Technology, *Communications of the ACM* 42, no. 5 (1999): 39–44.
44. See Rieh and Hilligoss, College Students' Credibility Judgments, and S. Shyam Sundar, The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility, this volume.
45. Wolfgang Schweiger, Media Credibility—Experience or Image? A Survey on the Credibility of the World Wide Web in Germany in Comparison to Other Media, *European Journal of Communication* 15, no. 1 (2000): 41.
46. R. David Lankes, Trusting the Internet: New Approaches to Credibility Tools, this volume.
47. Eysenbach, this volume.
48. Frances Jacobson Harris, Challenges to Teaching Credibility Assessment in Contemporary Schooling, this volume.
49. Jim Giles, Internet Encyclopedias Go Head to Head, *Nature* 438 (December 15, 2005): 900–901.
50. See Harris, this volume.
51. Miriam J. Metzger, Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research, *Journal of the American Society for Information Science and Technology* 58, no. 10 (2007); Metzger et al., Credibility for the 21st Century.
52. David R. Danielson, Web Credibility, in *Encyclopedia of Human-Computer Interaction*, ed. Claude Ghaoui (Hersey, PA: Idea Group, 2005), 713–21; B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* (San Francisco: Morgan Johnson, 2003); Metzger et al., Credibility for the 21st Century; Rieh and Danielson, Credibility: A Multidisciplinary Framework.
53. It is interesting to note, though, that high cost certainly does not ensure credibility. For instance, "yellow journalism" of the late 1800s relied on partisanism, strong editorial opinions mixed with "news," and tactics of sensationalism. Similarly, contemporary publications like the *National Enquirer* remain wildly successful, in spite of (and partially because of) the nature of their stories and reportage, which are commonly understood not to be credible.
54. Rieh and Danielson, Credibility: A Multidisciplinary Framework.

55. Danielson, Web Credibility; Flanagin and Metzger, Perceptions of Internet information credibility; Flanagin and Metzger, The Role of Site Features; Joseph W. Janes and Louis B. Rosenfeld, Networked Information Retrieval and Organization: Issues and questions, *Journal of the American Society for Information Science and Technology* 47, no. 9 (1996): 711–15; Metzger et al., Credibility for the 21st Century; Rieh and Danielson, Credibility: A Multidisciplinary Framework.
56. Danielson, Web Credibility; John W. Fritch and Robert L. Cromwell, Delving Deeper into Evaluation: Exploring Cognitive Authority on the Internet, *Reference Services Review* 30, no. 3 (2002): 242–54; Metzger, Making Sense of Credibility on the Web.
57. See Sundar, The MAIN Model, this volume.
58. See Eysenbach, this volume.
59. Gunther Eysenbach and Christian Kohler, How Do Consumers Search for and Appraise Health Information on the World Wide Web? Qualitative Study Using Focus Groups, Usability Tests, and In-depth Interviews, *British Medical Journal* 324 (March 9, 2002): 573–77.
60. See Fred W. Weingarten, Credibility, Politics, and Public Policy, this volume, for a discussion of policy issues related to credibility.
61. Danielson, Web Credibility; Flanagin and Metzger, Perceptions of Internet Information Credibility; Rieh and Danielson, Credibility: A Multidisciplinary Framework.
62. Alexander and Tate, *Web Wisdom*.
63. Nicholas C. Burbules, Digital Texts and Future of Scholarly Writing and Publication, *Journal of Curriculum Studies* 30, no. 1 (1998): 105–24.
64. Rachna Dhamija, J. D. Tygar, and Marti Hearst, Why Phishing Works, *Proceedings of CHI 2006 Human Factors in Computing Systems*, Montreal, Quebec, 2006, 581–90.
65. See Lankes, this volume.
66. See both Lankes and Eysenbach, this volume, for discussions of these and related issues.
67. Nicholas C. Burbules, Paradoxes of the Web: The Ethical Dimensions of Credibility, *Library Trends* 49 (2001): 441–53.
68. As cited in Rieh and Danielson, Credibility: A Multidisciplinary Framework, 311.
69. Lenhart, Madden, and Hitlin, *Teens and Technology*.
70. Eysenbach, this volume.
71. Lenhart, Madden, and Hitlin, *Teens and Technology*.
72. Ibid.
73. E.g., Scott Carlson, The Deserted Library: As Students Work Online, Reading Rooms Empty, Leading Some Campuses to Add Starbucks, *Chronicle of Higher Education*, 2001, <http://chronicle.com/free/v48/i12/12a03501.htm> (retrieved August 5, 2005).
74. Charles R. Martell, The Ubiquitous User: A Reexamination of Carlson's Deserted Library, *Portal: Libraries and the Academy* 5, no. 4 (2005): 441–53.
75. Lenhart, Madden, and Hitlin, *Teens and Technology*; Susannah Fox and Mary Madden, *Generations Online*, Pew Internet & American Life, 2005, http://www.pewinternet.org/PPF/r/170/report_display.asp (retrieved July 24, 2006); Rieh and Hilligoss, College Students' Credibility Judgments; Teenage Research Unlimited, *Teen Internet Safety Survey*, National Center for Missing and Exploited Children and Cox Communications, 2006, <http://www.netsmartz.org/safety/statistics.htm> (retrieved July 24, 2006).

76. Lenhart, Madden, and Hitlin, *Teens and Technology*.
77. Eastin, this volume.
78. See *ibid.*
79. See Lankes, this volume, and Eysenbach, this volume, for similar arguments.
80. The following section was informed by online dialogues with researchers, educators, and interested members of the public, as well as by youth participants in the Global Kids Digital Media Essay Contest. Over thirty people participated in the online dialogues, which were held in the fall of 2006 and were sponsored by the MacArthur Foundation as part of the Digital Media and Learning initiative. Excerpts of the online discussions may be found at <http://spotlight.macfound.org/main/category/c/Credibility/>. For information about the Global Kids Digital Media Essay Contest, including the winning essays, see <http://www.globalkids.org/>.
81. See Lankes, this volume.
82. See Weingarten, this volume, for a similar discussion with regard to policy.
83. See Harris, this volume.
84. See *ibid.*, for a full discussion of these educational issues.
85. Denise E. Agosto, A Model of Young People's Decision-Making in Using the Web, *Library & Information Science Research* 24 (2002): 311–41; Denise E. Agosto, Bounded Rationality and Satisficing in Young People's Web-based Decision Making, *Journal of the American Society of Information Science and Technology* 53 (2002): 16–27; Raya Fidel, Rachel K. Davies, Mary H. Douglass, Jenny K. Holder, Carla J. Hopkins, Elisabeth J. Kushner, Bryan K. Miyagishima, and Christina D. Toney, A Visit to the Information Mall: Web Searching Behavior of High School Students, *Journal of the American Society of Information Science* 50 (1999): 24–37; Sandra Hirsh, Domain Knowledge and Children's Search Behavior, in *Youth Information-seeking Behavior: Theories, Models, and Issues*, eds. Mary K. Chelton and Colleen Cool (Lanham, MD: Scarecrow Press, 2004): 241–70; Andrew Large, Information Seeking on the Web by Elementary School Students, in *Youth Information-Seeking Behavior*, 293–320.
86. See Rieh and Hilligoss, College Students' Credibility Judgments, this volume, for a discussion of when youth are more likely to invest effort in assessing credibility.
87. Metzger et al., Credibility for the 21st Century.
88. E.g., Metzger, Making Sense of Credibility on the Web; Rieh and Hilligoss, College Students' Credibility Judgments, this volume.
89. See Sundar and Eysenbach chapters, this volume.
90. Eastin, this volume.
91. See *ibid.*
92. Rieh and Hilligoss, College Students' Credibility Judgments, this volume.
93. Sundar, The MAIN Model.
94. Eysenbach, this volume
95. Sundar, The MAIN Model.
96. Rieh and Hilligoss, College Students' Credibility Judgments, this volume.
97. Flanagin and Metzger, Perceptions of Internet Information Credibility; J. Ha, Questioning Internet Credibility: A Test on the Cyber Marketplace (paper presented at the Annual Meetings of the

Digital Media and Youth

27

International Communication Association, Washington DC, May 2001); Johnson and Kaye, *Cruising Is Believing?*

98. See Lankes, *Trusting the Internet*, this volume.

99. Eastin, this volume.

100. Rainie, *Life Online*.

101. Harris, *Challenges to Teaching Credibility Assessment in Contemporary Schooling*.

102. H. Jenkins, *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century* (white paper produced for the MacArthur Foundation's Digital Media and Learning initiative, November 13, 2006), <http://www.digitalllearning.macfound.org> (retrieved November 14, 2006).

103. Bennett, *Civic Life Online*.

104. Harris, *Challenges to Teaching Credibility Assessment in Contemporary Schooling*.

105. Weingarten, *Credibility, Politics, and Public Policy*.

iSCHOOLS PRESENT THE THIRD ANNUAL

iConference 2008 :: iFutures: Systems, Selves, Society

February 28 - March 1, :: UCLA :: Los Angeles, CA

Citation: Lankes, R. David. "Trusting the Internet: New Approaches to Credibility Tools." *Digital Media, Youth, and Credibility*. Edited by Miriam J. Metzger and Andrew J. Flanagin. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. Cambridge, MA: The MIT Press, 2008. 101–122. doi: 10.1162/dmal.9780262562324.101

Copyright: © 2008 Massachusetts Institute of Technology. Published under Creative Commons Attribution-Noncommercial-No Derivative Works Unported 3.0 license.

Trusting the Internet: New Approaches to Credibility Tools

R. David Lankes

Syracuse University, School of Information Studies

It has been said that the Stone Age did not end because humans ran out of stones. Instead, Stone Age technology was superseded by new tools and capabilities. At some point in history, it simply became more advantageous to adopt new methods and tools rather than trying to solve problems inherent in older methods. Society may soon be at this inflection point in terms of how people, and particularly youth, identify credible information, abandoning traditional methods of determining credibility that are based on authority and hierarchy for digital tools and new network approaches. Far from being a negative development, new methods and tools for determining credibility may reflect a more distributed and open approach than in the past. Such an approach has important implications for how youth are educated, how policy is determined, and how future information systems are built.

This chapter first highlights some reasons why youth, the institutions that serve them, and society as a whole are moving online, as well as some of the consequences of this move—namely, the paradox of “information self-sufficiency.” A reformulated vision of credibility is offered in this context, which highlights features of digital information and networks. Then, a shift among credibility tools and techniques from traditional authority models to more of a “reliability approach” is discussed. Based on this, a framework for understanding the implications of information self-sufficiency for learning in a networked digital world is presented. This framework is used to highlight the often invisible effects that technology has upon credibility. Finally, implications of this are explored and current and anticipated developments on the Internet are considered. The chapter concludes by discussing implications of the information self-sufficiency paradox in the context of the education of youth in the current digital media environment.

The Shift Toward Increased Information Self-Sufficiency

There is little doubt that in the United States and other developed countries, citizens are increasingly relying on the Internet to gather information. Seventy-three percent of U.S. adults are Internet users, and 42 percent of Americans (about 84 million) now have broadband connections at home, up from 29 percent in January 2005.¹ The numbers are even more striking for youth. Lenhart, Madden, and Hitlin recently reported that 87 percent of youth in the United States ages 12 to 17 are Internet users, 51 percent of whom say that they use the Internet on a daily basis.² Other estimates indicate that half of children in grades 1–5 are online, as are 80 percent of high school students.³

Increasing Internet reliance is also evidenced by the dramatic increase of self-service options available to Internet users. Today, individuals are expected to book their own airline tickets, determine their own retirement plans, and even decide between life-and-death medical treatments using Internet tools and information available on the Web, without the assistance of traditional information intermediaries.⁴ However, the quality of these services varies. Examples of bad self-service options abound and nearly everyone has a horror story of getting tangled in a phone tree or diving deep into a company Web site desperately looking for a phone number to call or person to e-mail. Yet, there are also plenty of examples where online self-support systems and customer service have been designed and implemented well. Examples include Lands' End live chat support service⁵ and package tracking through UPS, FedEx, and other overnight carriers, to name a few.

The trend toward online self-support services and "disintermediation" more generally⁶ is also apparent in nonprofit and governmental sectors. For example, the National Science Foundation has supported extensive research into "digital government" that seeks to provide support of digital integration over a wide variety of government tasks such as electronic voting, public comments, security, and more.⁷ In the K-12 arena, Rice discusses several national policy initiatives involving the use of digital media to expand educational opportunities for U.S. students.⁸

Youth in particular are increasingly dependant upon online systems and support. From homework help services such as Tutor.com to completely virtual high schools,⁹ school-based learning is increasingly complimented by online services. As early as 2001, for example, Lenhart, Simon, and Graziano noted "71% of online teens say that they used the Internet as the major source for their most recent major school project or report."¹⁰ This increasing use and reliance on the Internet means that for a growing percentage of students the quality of online services and self-support options can directly affect their learning.¹¹

Presumably, the chief advantage to customers of online information self-sufficiency is greater control and satisfaction. However, this may not be the reality. ServiceXRG found that of the 60 percent of customers who used Web self-service, only 23 percent reported that they found what they were looking for online,¹² suggesting that only a small minority who choose to use online self-service features are able to find content that satisfies their needs. Thus, and ironically, many customers turned online actually require *additional* customer support. Unfortunately, such outcomes are not atypical.¹³

This shift to digital over physical media, however, is only partly a response to citizen demand for increased self-service. Another important factor is economic pressure felt by organizations to minimize cost by eliminating expenses such as printing and customer service personnel. According to the Web consulting firm Adaptive Path, in the last few years companies have "pushed labor-intensive tasks out to the customer, and they did so in a way that provided the customers with direct access to and control over information that they care about."¹⁴ The advantage to organizations is substantial cost savings. For example, ServiceXRG, a market research firm, found that whereas first contact closure cost of phone transactions is \$49.10, it is \$36.70 when done via e-mail, and only \$11.60 via Web self-service.¹⁵

The awkwardness with which some organizations have shifted the responsibility of support from employees to users is understandable given that the transition to digital methods of information creation, storage, analysis, and distribution has happened in an astoundingly short period of time. In less than fourteen years, the U.S. federal government, for example, has gone from mandating executive agency Web sites (Initially little more than simple online

brochures), to requiring electronic voting,¹⁶ Web-based submittals of grant applications,¹⁷ and electronic banking transactions. There simply has been too little time and too much change in the Internet environment to formalize and codify "good online service." To put it bluntly, what users need in order to take charge of their own online decision making is at best an art and, more often than not, a series of trial-and-error solutions.

Moreover, what may be called "information self-sufficiency" or "disintermediation" is far from an Internet-only phenomenon. Anyone who has checked into a flight at the airport knows that the number of touch screens is steeply on the rise, while ticket agents are in steep decline. Libraries now have self-checkout kiosks as do grocery and other retail stores. Information self-sufficiency is obvious in a world where so many aspects of life are either online (e.g., online banking, music downloads, shopping, medical information, government documents access) or are facilitated by online transactions (e.g., electronic fund transfers, traffic management systems, automated payroll systems). In the end, information self-sufficiency has dramatic impact: it affects how products are marketed, how organizations manage information, how courts assess liability, and even how the current and future workforce is trained.

Information Self-Sufficiency and the Heightened Importance of Credibility

Although information self-sufficiency could (and should) be examined in a number of ways—including economic, political, and even in terms of social and class roles—this chapter concentrates on the effects of information self-sufficiency on credibility, or the believability (i.e., the trustworthiness and expertise) of some source of information. Indeed, credibility, particularly in the context of youth and the Internet, is fundamental to the long-term success of information self-sufficiency and, by implication, the success of digital media themselves. An emphasis on credibility is crucial in the success of society's digital migration because information is increasingly disconnected from its physical origin and, as a consequence, the credibility of information has taken on new complexities, with new implications.

Consider the simple act of buying a book over the Internet. When people buy a book online, they are not basing their buying decision on a single, physical item. They are instead basing the buying decision on information *about* a book (e.g., its price, author, shipping terms, reviews). Even if they had previously checked the book out of a library, or paged through it at the local bookstore (where they could have bought the book based on the information it contained *and* the physical attributes of the book such as whether it is in good shape, etc.), in the online environment they are simply putting in an order for a book based on some proxy, such as a digital image of a physical item.

This seemingly mundane and obvious fact actually has sweeping implications for credibility on the Internet. In particular, for any online transaction that involves delivery of some physical item, a central means of determining credibility—physical examination and testing—is gone, or is at least greatly diminished (since one can usually return the book after a physical inspection at the end of the process). No more is it possible to review a signature to make sure it is original. No more is examination of the quality of paper used in a publication practical. No more can one audit a physical inventory prior to purchase.

Another example that serves to illustrate the point is a digital fingerprinting system for use by law enforcement personnel in New York State. When suspects are arrested, instead of inking fingers and making cards with copies of fingerprints, in many jurisdictions prisoners now place their fingers on glass and have them digitized. The system was intended to speed

up processing of fingerprints, and to make them more accessible for searches (P. Lorenzo, personal communication). However, removing the physical collection of fingerprints had an unintended consequence. At one arrest a processing officer took a suspect's name and date of birth, and simply copied a fingerprint already stored under that name and date of birth into the new case file (which was against policy). It turns out the date of birth and name given by the suspect were false, and the prints copied belonged to a prisoner already incarcerated. More striking, when the mistake was revealed, and the suspect's prints were actually scanned, it was discovered that he was wanted for another crime—murder. This mistake was possible, and undetectable, in a digital networked environment because every transaction is an information-only event: that is, transactions exist only as digital information with no physical record or artifact. As such, some transactions work better in the digital environment than others.

For example, while buying books online has become popular, buying homes online is still a rarity. This is due both to the size of the transaction, and also to how much of an object's information is intrinsic and how much is extrinsic to its physical form.¹⁸ The more mass produced and standardized an item is, the more information can be "separated out" from that item. In the case of the book, nearly all of the information within the book—including the author, the price, the title, and even the visual images of the pages themselves—can be recreated accurately in digital form. Thus, the information is extrinsic to the physical item. With a house, however, the condition of the roof, the true color of the paint, the "feel" of the neighborhood, and the creakiness of the floors are all intrinsic qualities that need physical inspection to assess accurately. This also explains why buying a new car online is not only possible, but is a growing industry. Although it is a large transaction, new cars are fairly standard, so a test drive of a car in New York is assumed to be identical to test driving the same make and model in California. Therefore, buying that new car in California does not require a second test drive. The same cannot be said of a house. Houses are unique, and so the information relevant to a buying decision is intrinsic to the house.

Since information is the only evidence available to people when making an online transaction (e.g., buying something, talking to someone, learning about something), the credibility of that information is essential. Moreover, methods of building trust, and methods to test assertions have changed. For example, if one goes to buy a book online, that person can no longer test the book to see if it is complete and in good condition prior to delivery. They must now trust the online vendor of that book. This is the great paradox in information self-sufficiency on the Internet: *end users are becoming more responsible for making information determinations, but because they have fewer physical cues to work with, they are becoming more dependent on the information provided to them by others.*

This paradox is complicated further by the fact that people are simply unable to, or fail to, recognize many of the more technical influences on the information with which they are provided in the first place. In fact, there is a great deal of information manipulation that occurs that is *never* perceptible to the user. Built into the tools themselves are filters, assumptions, biases, and outright distortions that can never be factored into a user's credibility decision.¹⁹ Indeed, there has been much discussion and research of how the media used to access information can affect perceptions of the credibility of that information.²⁰ There have also been a number of studies into how the online environment itself affects credibility.²¹ Together, this work highlights how credibility can be both determined and manipulated by technical elements, such as load time of Web pages or site design.²² Research also points out that while technology influences credibility decisions, it is often invisible to the end user.²³

Perhaps because these things are so invisible to most people, this point is completely missed in many examinations of how users make credibility decisions online.

Youth in particular are increasingly dependent on software and hardware tools while accessing information and services on their own within the digital environment. Put simply, there is no way to be on the Internet without some intermediating piece of technology. Be it a Web browser, a cell phone, or some other tool, information flowing to and from the Internet, as well as information crucial for determining credibility, all flow through seemingly invisible agents of code and silicon. The Lankes/Eisenberg Architecture²⁴ can be used to highlight the kind of information manipulation that occurs solely in the province of the tools of the digital environment. This architecture divides the Internet and, by extension, digital networks into four distinct layers: infrastructure, application, information service, and user. Each of these layers can manipulate information in a way that is completely transparent to the recipient. Consequently, there are implications of each layer in terms of youth learning.

Infrastructure is composed of hardware (e.g., routers, protocols) used to move information from one place to another on the Internet, and the organizations, such as Internet Service Providers (ISPs), that provide and maintain these mechanisms. This layer is often the most invisible to end users, yet can have a profound impact on the information available to users to enable them to make credibility assessments. For instance, many people do not realize that infrastructure providers can easily block traffic to and from certain destinations, and can make such blocked traffic invisible. When a school blocks access to certain Web sites, they may post a message to a student's browser stating that the site is blocked, or they can more simply provide a "site not found" indication to a user's browser, the same error it would send if the user misspelled a URL. One component of people's credibility assessments in the online environment is the comprehensiveness of the information they obtain.²⁵ As such, filtering programs used in schools may negatively influence students' perceptions of the comprehensiveness, and perhaps even the impartiality, of Internet-based information. In addition, ISPs can block access to any application, disabling software such as instant messaging or social networking at the network layer. The user, not aware of such a block, would only know that their IM (instant messaging) program did not connect to a server and may assume that the error lies in the remote server, thus erroneously affecting their credibility assessment of the remote server, rather than the infrastructure provider.

Applications on the Internet consist of software that allows information to be exchanged between different actors on the network. Applications include Web browsers and instant messaging clients, as well as high-level protocols such as HTTP that transfer Web pages. This broad category covers everything from e-mail applications that automatically mark incoming messages as "junk mail" to the Simple Mail Transfer Protocol (SMTP) that enables e-mail over the Internet, including spam. Spam filters are excellent examples of technology affecting credibility in a nearly invisible way. Many schools have implemented spam filters based on opaque and often proprietary algorithms at the organization level, discarding numerous e-mail messages before any human eyes ever see them. While this type of protection from inappropriate or harmful content can be good, it can also have the unintended consequence of inadequately preparing youth for the digital world they may encounter outside the school environment. Indeed, as both Harris²⁶ and Weingarten²⁷ argue, blocking young people's exposure to information that may not be credible is probably not the best strategy for teaching students to identify and defend themselves against such information, which they will almost certainly encounter at some point in their lives.

Information services are organizations that use applications and infrastructure to meet users' informational needs on the Internet, such as Google and MySpace. From a credibility standpoint, there are ample studies that look at how information services such as Google skew results in their search engines.²⁸ In fact, most search engines, including Google, determine the "quality" and "relevance" of sites using a "link-popularity" metric. This metric selects which Web sites to display and the order in which to display them on the search results page based on how many other sites link to a site. Consequently, more popular pages are selected and are displayed higher in the search results. Because few people go beyond the first few pages of the search output, however, "even if a page is of high quality, the page may be completely ignored by Web users simply because its current popularity is very low."²⁹ This kind of a system sets up a sort of "popularity equals credibility" heuristic³⁰ that could be dangerous or at least disadvantageous to students' learning. As another example of the bias inherent at the level of information services, and the resulting credibility implications, top results tend toward shopping and technology services in Google.³¹ Without knowing this, youth may assume that top results are the "best" regardless of context.

Finally, the *user layer* is composed of individuals and groups, such as teachers and students, who primarily seek and consume information on the Internet to meet their own information needs. Of course, users bring their own biases to both information consumption and production that may affect their credibility perceptions as well as learning. Perhaps the best example is the online encyclopedia Wikipedia. While the collaborative editing afforded by Wikipedia may in fact produce credible information, users must be cognizant that contributors and their contributions may be biased, uninformed, or outdated and should evaluate the information accordingly. A young person using Wikipedia for class assignments who is unaware of the process and accompanying pitfalls of collective editing is thus vulnerable to relying on misinformation.

Decisions at each of these layers can affect credibility judgments by all Internet users, but perhaps particularly by youth, given their heavy reliance on digital media, coupled with meaningful developmental and experiential differences between youth and adults.³² In the end, having tools as intermediaries deepens the information self-sufficiency paradox by making youth more independent in their information seeking, while simultaneously making them more dependent not only on the information they receive but on the tools they use to access this information.

Shifting Credibility from Authority to Reliability

One outcome of this paradox is a shift from an authority-based approach to credibility to a "reliability approach." This shift represents a sea change from the way in which credibility has traditionally been conceived. Whereas credibility has historically been tied to concepts of authority and hierarchy, in the reliability approach, users determine credibility by synthesizing multiple sources of credibility judgments.³³ Both the need for synthesis and the richer set of resources to be synthesized are products of the pressure for participation enabled and imposed by networked digital media.

Traditional approaches to credibility strongly emphasize authority, where a trusted source is used to inform an individual's credibility determinations. In essence, trusted sources are used to "vouch" for the credibility of a given piece of information. People may have many such authority sources, and may themselves serve as an authority in various settings. The

process of becoming or choosing an authority is a process of developing trust and seeking coherence and consistency in the authority.

Some claim that authority is dead on the Internet, for example, some have said Wikipedia (http://en.wikipedia.org/wiki/Main_Page) and group editing have become at least as authoritative as traditional methods for producing encyclopedias, if not more so.³⁴ Others feel that blogging will supersede or at least parallel the authority of traditional news outlets.³⁵ There are indeed a number of ways in which traditional means of authority have been supplanted by open, flat structures for information creation and credentialing. However, to call this tantamount to the death of authority is, at best, an imprecise use of terminology. This new paradigm is not without authority, but more sophisticated methodologies may be required for evaluating it.³⁶ Moreover, while they may not have been as celebrated or accessible as they are currently, history is replete with examples of waves of centralization and decentralization of authority. The dramatic increase in information self-sufficiency has merely led to celebrating the large-scale nature of this decentralization today.

Beyond this historical view, most people use the term *authority* to refer to a single entity, or small set of entities. That is, authority has traditionally been conceived of as hierarchical and centralized. For example, libraries have come together to invest the power of authority in a given institution. For example, the Library of Congress keeps a file of book authors' names, birth, and death information. When citing an author (or including them in a library record) someone can consult these authority files for proper spelling, aliases, and confirmation of an author's identity. This can be thought of as "authority by consent," where all parties in a situation or context agree on who is the authority—that is, who provides the accurate information.

Nonetheless, it might be more precise to use the terms *authoritarian* and *authoritative* to clarify traditional notions of authority. *Authoritarian* is defined as the enforcement of an authority: in essence, the removal of choice by force of law, policy, structure, or some other means. *Authoritative*, on the other hand, is authority granted on the basis of perceptions of trust and expertise. The former is active and enforced, the latter is earned. Wikipedia is more likely to be the death of an authoritarian view that encyclopedias come from only Britannica, Groliers, or other large publishing houses, than it is to be the death of authority per se.³⁷ Thus, the problem of determining the credibility of Internet-based information is not a crisis of authority, but rather a crisis of choice. There are simply currently more choices about whom to trust. Although this is true for virtually all media venues to some degree, the scale of choice makes the Internet particularly affected by shifts in authority.

Libraries have been wrestling with this issue of shifting modes of authority since at least the early 1990s. Many in the library community—though not all, of course—have reacted to the overwhelming number and variety of information choices available to their clients via the Internet by adopting an authoritarian view that the library is where to get good, accurate, and credible information, while the open Internet is filled with bad and wrong information. One famed tag line of the time was, "Information on the Internet is free, but you get what you pay for." Many want the library to become a preferred provider of information, yet the concept of "preferred" only works in an authoritarian view when there is someone who can make others prefer or select something over something else.

This resistance to the democratization of authority among librarians is ironic because there are few professionals better suited to the authoritative world of the Internet than are librarians. They have a culture of open and free expression and access to ideas. They are generalists who move agilely across different topical domains. They are skilled at searching

out information, and locating potential biases in information. Their enterprise (i.e., the library) has little invested in the production of information, and much invested in the consumption of information products from a wide variety of sources. Furthermore, librarians already have a reputation as authoritative, not authoritarian.

The resistance to the democratization of authority is not limited to librarians. Teachers, college professors, doctors,³⁸ and indeed just about any information intermediary has had to wrestle with the new environment of plurality in authority. For some areas of the economy, the effects have been drastic. For example, the travel industry has been rocked by the growth in online bookings.³⁹ And, although people may not be buying houses online, they are using Internet real estate listing services to even the marketplace.⁴⁰ Perhaps one of the most striking examples of the role of authority can be seen in the insurance industry. As Levitt and Dubner note, the simple act of allowing consumers to directly compare term life insurance rates from different organizations over the Internet, rather than depending on the authority of a life insurance salesperson, led to the cost of term life insurance in the United States to drop by \$1 billion.⁴¹ Interestingly, in this case the concept of an authority in a field (an insurance agent) was actually used to mask truly credible information (in the form of lower prices).

The decentralization of authority has turned out to be particularly attractive for teenagers. This shift against stated authority and hierarchy on the Internet is perfectly matched to teenagers' own internal shift against authority of all types.⁴² Note, for example, that the most common tools used by teens include IM, MySpace, and Google, all of which allow for a sort of level playing field, where the "right" or authoritarian answers are virtually indistinguishable. IM, a peer-to-peer technology that allows for instant access to friends and social peers, may constitute a wide network of potential experts in this environment built by the teen, not imposed by some adult "expert." In other words, the technology allows greater opportunity for young people to themselves become authoritative experts in many areas, potentially even shifting the power balance between children and adults in some cases. Eysenbach makes a parallel argument for how digital media have impacted the relationship between doctors and patients.⁴³

Not surprisingly, many adults feel it is a problem that teens may not look to vetted and traditional sources of "valid" information. Their answer to the problem is often phrased in educational terms: "If only we show them the good stuff, they'll use it." This approach has several guises, normally in terms of literacy: digital literacy, information literacy, information problem solving, and those approaches that rely upon checklists to determine "right" and "wrong" information. However, while these approaches can have a positive effect in certain populations, such as undergraduates and elementary school children, under certain conditions they do not always "fix" the problem.⁴⁴

Of course, the concept of multiple authorities existed prior to the Internet. But, the cost in terms of time, money, and even reputation to seek out a multitude of authorities was very high. With the Internet and other digital media, the range of possible authorities has expanded greatly. Before the Internet, for example, a young person might have to rely on his or her local television news or newspaper to predict the weather. Now he or she can go to the Weather Channel, AccuWeather, or even the National Oceanic & Atmospheric Administration for weather information. The task then becomes determining from a number of sources which among them is most credible.

Moreover, a consequence of the culture of information self-sufficiency is that people, even youth, can more easily become authorities themselves. Not only do users have more

sources to choose from, but now they can also access data and information directly, thereby avoiding traditional authority sources altogether.⁴⁵ For example, now young people can gather weather data directly from a variety of satellites and radar installations (including home-based weather stations), and train themselves, until they feel they have sufficient expertise and trustworthiness to credibly interpret information. As users take it upon themselves to become authorities by directly evaluating, synthesizing, and even producing information themselves, the notion of a singular authority ends, and "reliability" becomes the predominant form of credibility assessment.

Reliability commonly refers to something or someone perceived as dependable and consistent in quality. If you have a reliable car, it is one that runs well over time. Reliability to the scientist is simply the consistency of data, such that the same treatment (e.g., questions, experiments, or applications) yields the same result over time. If an authority approach is exemplified by believing that a given news anchor will give a credible answer, then switching from news station to news station looking for commonalities in the same story exemplifies a reliability approach. Reliability approaches to credibility can be seen in the physical world. The fact that a person's signature is seen as a marker of credibility in legal settings is the belief that a person signs his or her name in a reliable fashion. Reliability is also seen in the cornerstone of good journalism: to confirm information with a second source.

Authority and reliability approaches are often used in conjunction in both the physical and digital world. In the aftermath of Hurricane Katrina in New Orleans in 2005, many residents turned to chat rooms and community-run Web sites to resolve contradictory and often false information coming from traditional channels and sources (including the federal government and the mass media). Local Web sites, such as NOLA.com, allowed communities to come together and share information. Users were able to hear from multiple sources, including eyewitnesses and residents, and to get a more accurate, complete, and thus credible picture of the situation in the neighborhoods. Users who provided the most consistently accurate information became trusted authorities, and traditional authorities (the government and the mass media) were deemed less credible than the users. This is but one example of how digital media have turned credibility on its head.

Perhaps the most common way to become an authority, however, is through reliability. If someone consistently gives out testable and *accurate* information, they are often seen as an authority.⁴⁶ Experts are not simply people with the most experience, but people who have the longest track record in using their experience successfully. The true power of reliability, however, is not only in its capacity to create authority but in its power to destroy it as well. While the necessary degree of reliable performance for authority status varies across contexts (e.g., a baseball player who bats .350 is considered outstanding even though he hits the ball less than half of the time, but a mathematician would not be considered an expert if she got math answers right only half of the time), it is clear that agents giving out unreliable information over time will lose their authority status and, by extension, their credibility.

Authority and reliability also have "halo effects," meaning that a person who is seen as an authority in one area may be presumed to be an authority in other domains as well, even without proven performance. This can be seen, for example, in celebrity product endorsements. Likewise, unreliable performance in one area can cast doubt on performance in other areas. Like an accountant who cheats at Monopoly, or a minister who cheats on his wife, loss of credibility in one area can cast suspicion on one's other domains of authority. These halo effects become even more pronounced in the digital environment when the information

about and from an agent (person, organization process) can be more voluminous, diverse, and mobile than in offline environments.

Ultimately, reliability and authority anchor the endpoints of a continuum of approaches to credibility. With authority, preexisting agreements are in place and assumed: the conversation is over. With reliability, by contrast, the conversation is open and ongoing. Networked digital media enhance this conversation and thus are more likely to encourage users to seek out more information and other people to come to a credibility judgment compared to traditional media. Since digital networks bring to youth an overwhelming amount of information and people to engage with in arriving at a credibility judgment, there is pressure for the tools they use to incorporate some ability to participate, or engage in a conversation with people and sources. Builders of digital network tools and information services have begun to respond by being more open and by creating more opportunity for user participation in content creation and dissemination. In so doing, the tools built for users today to find and use credible information facilitate reliability approaches to credibility assessment, and thus learning via constant "conversation."

Digital Media and Credibility by Reliability

Recent trends in digital media facilitate the shift to a reliability approach of credibility assessment. The Internet is by its very design open, providing only minimal control at the network level. The guiding technical protocol of the Internet, Transmission Control Protocol/Internet Protocol (TCP/IP), simply breaks data into packets and makes sure these packets get to the proper destination with fidelity. What these data packets contain, what they do, and how they do it is completely ignored by most users. From this very simple packet-switching technology, Internet users have built e-mail, the Web, instant messaging, and all of the services that users enjoy today. Indeed, the Internet's interoperability, open access, and decentralized control are especially well suited to innovation on the scale of the individual user.⁴⁷

As a result, even higher-level Internet functions are not defined or controlled by the Internet itself. Indeed, there is no central authority that controls the Internet, beyond its basics such as domain name registration and the structure of TCP/IP. Even so-called governing bodies such as the World Wide Web Consortium (W3C) merely suggest standards that are often embellished, altered, or ignored by software developers.⁴⁸ The Internet is, in a very real sense, an agreement, and an ongoing conversation where organizations and individuals share information.

This open architecture and decentralized control provide enormous capacity for participation, which contributes to accurate credibility assessment through reliability among multiple shared perspectives and experiences. Yet, there is nothing inherent in digital networks that makes infrastructure participatory. There are many historical examples of large-scale networks that were centrally controlled, where infrastructures were provided to users with little input by those users. Commercial online networks ranging from CompuServe, Prodigy, and America OnLine to not-for-profit Freenets and BitNet networks were "black boxes" that only allowed users to manipulate pre-made tools. These are historical examples, however, because, even in the case of AOL, these proprietary networks have either been superseded by the Internet, or have had to radically change their underlying infrastructures to accommodate the Internet.⁴⁹ Thus, although involvement in the infrastructure and at the level of specific software applications is not a certainty, the Internet's evolution appears to favor involvement and participation.

Indeed, involvement and participation have led many Internet software designers to view the very process of software development as a means toward credibility. For instance, so-called open source software at its most basic is when the developer of a piece of software makes the underlying source code of the software application available to the public. Anyone with sufficient programming skills can then take the source code and analyze it, add to it, or incorporate it into another software package. On the Internet, this simple concept has been expanded to a more complex approach to any system development.⁵⁰ While anyone with sufficient skill can use open source software, if they improve or expand the original code, they must give the additions back to the open source community.

Open source software and the associated movement propose a new concept of credibility in terms of tools: credible tools are ones that are built in the open, where a conversation on the merits and structure of infrastructure can be debated and tested. This stands in stark contrast to more traditional models of credible software development. In a traditional approach to software development, credibility is defined by the organization that produced the software. This organization would have some standing in the community, good practices, and a proven track record of quality products (i.e., authority). Interestingly, this is often proffered in terms of security. A system is "secure" if few have access to its inner workings, and if the few programmers who put it together used a common quality process. The open source approach takes an opposite stance: To be secure, software must first be transparent and then tested by others (i.e., be reliable). The idea is that, if everyone can see the code, and can test it, flaws will be easier to find and because the process of implementation is tested in public, everyone can trust the product. Open source advocates would argue that by being able to dissect and test all aspects of a piece of software, right down to the very source code, they can better determine both the trustworthiness of a tool as well as the expertise of the tool's creators (i.e., its credibility).

While there is an ongoing debate between the "open" and "closed" software development approaches, there is no question that it has changed the shape of tool building on the Internet. For example, the Internet produced the concept of "open beta" where software products were made available to any Internet user to test while still in production. Over time, these beta test processes run longer, sometimes never leaving beta (i.e., "continuous beta"). Even traditional software developers such as Microsoft and Adobe have moved toward a more open approach "to enlist a large army of bug testers to help iron out any kinks"⁵¹ in their products.

This openness goes a long way toward addressing the information self-sufficiency paradox among youth, but only for those with sufficient technology skills and education. That is, youth with sufficient technical skills now have the ability to choose and shape the tools themselves. Skilled youth who are wary of a given Web browser can use another one, or even write their own, often in concert with others sharing a similar concern. If a school wants to know exactly how its e-mail program is filtering spam, it can now shop multiple filtering packages until it finds the right one, sometimes evaluating the very source code of the filter.

This ability of *skilled* users fluent in the technologies of the Internet has important implications for youth and education. If society wants youth to be truly able to make credibility decisions in digital networks, then youth must understand the technical nature of the network itself—from the use of tools to the creation of tools. Simple use skills, such as browsing the Web, are insufficient to truly understand the role that tools play in the credibility of Internet-based information. Furthermore, if schools and other institutions prevent youth from participating in the underlying infrastructure,⁵² they are limiting youths' ability to

resolve the information self-sufficiency paradox and, by extension, limiting youths' ability to learn about and act upon credibility.

This limitation, and youths' reaction to limiting access to the underlying infrastructure of digital networks, can be seen in content filtering put in place at many schools and libraries today. When faced with limited access to the Web, many youth have begun to document the limitations of the filters themselves. Privacy organizations and civil liberties groups have joined students in challenging the efficacy of filters, and have called for change.⁵³ Because the infrastructure that these filters are built upon is open, and because the filter is not part of the network itself, it can be modified or replaced by those who possess the skills to do so. Suddenly the invisible nature of the network itself is visible—and indeed debatable.

The "Credibility Conversation"

To be effective both at constructing reliability-based credibility assessments and in working collaboratively to build innovative technical tools, youth must be able to engage in appropriate conversations with appropriate others via the Internet. Accordingly, Nielsen/NetRatings indicates that teens' usage of digital media is shifting. While the top sites for youth ages 12 to 17 used to be those offering a selection of instant messaging buddy icons, in the last three years the most popular sites have shifted to social networking sites or those providing assistance with social networking content, profiles, and page layouts.⁵⁴ While social networking sites such as MySpace, the Facebook, and others are relatively recent, teens' desire for online social interaction is not new. Teens have preferred social uses of the Internet for some time. For example, in 2001 Lenhart, Raine, and Lewis found that extremely popular uses of the Internet among teens included e-mail, instant messaging, and visiting chat rooms.⁵⁵ Teens' desire for social Internet experiences also shows up in institutional education settings, where Simon, Graziano, and Lenhart found that a large percentage of teens "say they use e-mail and instant messaging to contact teachers or classmates about schoolwork."⁵⁶

Beyond simply a person's age or cultural fads, the migration to social applications on the Internet is in part due to the very nature of learning. Conversation theory,⁵⁷ for example, proposes that learning and knowledge are gained through the interaction of agents around ideas as they go back and forth describing an idea. This interaction can then be used to develop new understandings and new knowledge through a process that contemporary learning theorists call *scaffolding*: one idea building upon another. Learning conversations can take place between two individuals (say a teacher and pupil), two organizations (negotiating a set of academic standards), or even two societies (debating over the best way to educate youth).

Framed by conversation theory, the utility of social interactions in online learning environments (be they formal or informal) for youth is obvious. As youth engage in conversations they learn, and the online environment allows for a greater scale and scope of conversants. Furthermore, as they encounter systems for learning in the digital environment, they may seek out tools to aid in interactions and conversation. The tools they seek out do not simply present credibility information, or third-party credibility assessments, but instead allow youth to participate in the conversation, and therefore the process of credibility verification and knowledge creation. Indeed, recent research suggests that members of open-source online communities are able to effectively build collective knowledge through "virtual re-experience," whereby individuals share their work using online tools to co-construct applicable knowledge.⁵⁸ Clearly, among other uses, youth are well poised to take advantage of precisely this type of conversation.

The concepts of openness, and indeed participation, in the development of tools exemplified in the open source movement can also be seen at the level of the Web sites and remote resources that users access on the Internet. Information services are under increasing pressure to open their sites and resources to youth participation. Information services today understand the power of participation. Where once services might count hits to a Web page, or unique visitors, today whole industries are devoted to analyzing a user's path through a Web site to discover what information he or she encountered, where this information led to successful "goal conversion," such as buying a product or, in an educational context, learning a piece of information, and where the information led to confusion or exiting a site. Educational organizations are now beginning to understand that the true power of the Internet for learning is not simply wide-scale and one-way distribution of information, but instead is getting closer to what is happening inside users' (learners', customers') brains. This trend follows larger forces at work in the educational setting that have resulted in the shift from accreditation and evaluative education bodies to outcomes evaluation and performance standards.⁵⁹

Recognition of the increasingly participatory nature of information services can also be seen in the rise of social networking sites, where the information service is little more than an infrastructure for user-contributed conversations and artifacts. Such services are the purest example of a trend finding its way into a large number of Web sites: users contributing to the ongoing conversation about an artifact (such as a book listed on Amazon), fact (such as an entry in Wikipedia), place (such as Flickr communities devoted to a given city), or other entity. In much the same way that traditional tool developers feel an increasing pressure for community input, content providers are also feeling an equal pressure for community input on their content. Part of this pressure comes from users who see participation as a crucial part of assessing the *credibility* of these artifacts, facts, and topics.

The advent of services that allow greater user involvement, such as blogs, social networks, and recommender, rating, or commenting in e-commerce and other sites, has led to increased demand by youth and other users for involvement in Internet services of all types. Key issues here in terms of credibility are, can one trust a person one has encountered online, and does that person know what he or she is talking about? Increasingly youth and others are looking to user-submitted comments, editorial reviews, and open conversations on a given topic, artifact, or idea to determine trust and expertise. And, in many cases where sites do not provide these functions, the youth community can take advantage of the open nature of the Internet to create their own. This has led to the rise of an entire class of network destinations known as "protest sites" devoted to user discontent with a given institution, person, brand, or product. In the physical world, youth might be limited in their protests to complaints to the principal or reliance on third parties such as better business bureaus, but online youth and others can seek out and build communities of discontented users regardless of geographic location.⁶⁰ Moreover, users new to the brand or item can now easily find such sites, allowing them to incorporate the dissenting views, as well as the official information, when making credibility decisions. In this way, youth can compensate for their relative lack of life experience⁶¹ by leveraging that of others. Of course, these user communities gain strength in numbers and are assessed in terms of their reliability, rather than their authority, as discussed earlier.

This style of grassroots organization has quickly spread from the consumer to the political arena. With an open network, opposing voices that challenge the credibility of some "official" information have equal access to the network and, at times, equal weight compared to official

information outlets and forums. This type of information "leveling" afforded by open digital networks can also be seen in schools. Youth, dissatisfied with either the information available to them from their classes, schools, and teachers, or with the venues for complaint, have turned to the digital environment to disseminate information of protest or information otherwise unavailable within the school. For example, there are now several sites rating teachers, professors, and classes. Blogs, MySpace, and other community sites are filled with "counter information" meant to provide greater context to, or to at least raise concerns about, the credibility of a given piece of information.

Of course, the Web also provides ample opportunity for users to post false information. There are many examples of "astroturfing," whereby marketing firms have set up seemingly grassroots sites that actually promote particular products or approaches.⁶² Certainly youth may put up false information on protest sites as well, either intentionally or unintentionally. The result is that credibility may be harder to determine for any piece of information. False information also increases pressure on information services to provide opportunities for community feedback, in essence inviting protestors and commentators into a controlled space where an organization can respond to counterinformation.

This pressure for a voice on the part of youth can be expected to result in more opportunities for discussion and conversation in both existing and emerging forms of digital media. It is not surprising that the distinctions between information seeking, learning, and communicating are breaking down. Where once users would e-mail (or instant message) in one application, and search the Web in another, they are now doing both simultaneously. This is not a simple matter of convenience, but rather is based on users' implicit or explicit preference for knowledge acquisition through conversation. Given what is happening on the Web in terms of social uses and applications, it appears that users are looking to talk with others in order to better evaluate what they find and to synthesize this information into actionable knowledge.

It is also likely that the trend of tracking users' conversations online will continue and will be enhanced. Information retrieval systems may well develop into conversational retrieval tools that link information by how others have used such information in their conversations. Just like libraries used to produce pathfinders and annotated bibliographies, users will soon be able to find a piece of information, such as a Web site, and follow that information to all of the other public information used in a given conversation. Such a holistic context will allow users to make credibility determinations about an item in the full context in which it resides. Digg.com provides an excellent present-day example of such a system.⁶³ When a Digg.com user finds an interesting news story on the Web, he or she can mark it, allowing other Digg users to not only view the site, but to discuss it in an online forum. What this does, in essence, is elevate the interest of one user to a group's consideration and evaluation.

Youth may understand this type of information seeking better than adults. Through e-mail, instant messaging, and texting, youth already create close-knit "research" teams that share findings and implicit credibility assessments.⁶⁴ Credibility in these contexts is not determined by the individual, or even the individual in interaction with a Web resource, but within a community engaged in a larger conversation. For school assignments, students may use sources they determine to be credible from their past experience, their heuristic appeal,⁶⁵ the input of experts such as teachers and librarians, and from others, including their peers. This "credibility conversation" may, however, extend past the actual assignment and well into the evaluation of that assignment. A low grade, or follow-up from teachers, may have an impact on how a student evaluates credible sources in the future, and how he or she conveys

information about the sources to others (via social networks or otherwise) who may want to use those sources in the future.⁶⁶ The grades and feedback that the peer group receives should also influence future credibility decisions. However, unless credibility is made explicit in this large and ongoing conversation, it will be ignored. That is, if teachers, parents, adults, and peers do not make credibility part of the discussion, and if there is no consistent result from either embracing or ignoring credibility, it will become invisible, not cognitively attended to by youth, and therefore not assimilated into knowledge. This situation is made worse by the growing educational environment that emphasizes "high-stakes" testing, which minimizes rich assessments and interactions between student and teacher.⁶⁷

The need to highlight credibility in conversations has implications for educational policy makers as well. If youth are exposed only to vetted and "safe" resources, often pruned of a great deal of context and conversation, how are students to gain the invaluable skills required to determine credibility on their own, outside of guided environments? Harris identifies such "conservative attitudes" as a structural challenge that must be overcome.⁶⁸ What policy makers must understand is that, as Weingarten⁶⁹ and Harris⁷⁰ argue, the unintended consequence of creating "safe" learning environments is an environment that limits learning about credibility.

The fact that technology constrains youth information before youth are able to make credibility decisions has startling implications for education. To begin, a common strategy to prepare youth to make informed credibility decisions is based around educational programs normally under some literacy rubric such as information literacy, digital literacy, or media literacy.⁷¹ These programs typically present young people with various strategies and checklists to determine the quality of information they find via digital media.⁷² While this approach makes sense for much of the information a student might access through digital networks, it does not make sense for all Web-based information. For example, how can one judge the credibility of information at a given URL when the true URL is masked by inline frames that show content from some third-party destination as if it were just a part of a normal page? Although this is a specific feature of the Web today, there are sure to be many analogs in future digital technologies. Moreover, there is no amount of literacy instruction that can prepare the average young person for the effects of a network infrastructure that they cannot control directly.

Conclusion and Recommendations

This chapter described the growth of information self-sufficiency and the related paradox where users are simultaneously more responsible for decisions, while also more dependent on information provided by others and the tools used to manipulate information. It discussed how the Internet and digital networks that allow for community participation in the construction of the underlying network and infrastructure have become the predominant model for existing and future digital media, and how this has led to credibility assessment by reliability rather than by authority. It presented a model in which knowledge is gained through conversation and reflection, and argued that digital networks and tools must address this reality. Finally, it showed how this need for participation and involvement has influenced information services, and how users themselves define and assess credible information.

Primary aims of the chapter have been to highlight trends that will most likely endure, continue, or expand as future digital media develop, and to show how youth are both affecting and being affected by these trends. Large-scale digital networks have extended

youths' ability to build their own social networks to aid in the assessment of credibility. They have also allowed youth to participate in the networks in meaningful ways, in essence elevating their own personal credibility in certain domains. However, this new reality has serious implications for youth, as well as for society as a whole.

To prepare youth to make fully informed credibility decisions, they must become fluent in the tools that facilitate the conversation and become aware of potential biases in the network technology itself. As noted by Harris, schools may be an ideal place to do this, but they are limited in their ability to do so.⁷³ Without technical fluency, however, students become dependent on, and often unaware of, the stakeholders who control the network's infrastructure and the policies they create.⁷⁴ Yet, as others in this volume have pointed out, the extent to which youth are involved shapes the Internet and the services available. The omnipresent nature of the digital environment is necessitating more technical fluency and greater explicit consideration of credibility and technology at an earlier age.

So, what are the implications for youth and credibility in the digital environment? The first is that youth, with enough training, now have equal opportunity to access the infrastructure they increasingly depend on. Unlike previous media technology widely adopted by youth, such as radio, television, and the phone system, youth can not only adopt technology, they can shape it at its most fundamental levels—at the level of infrastructure. Furthermore, with the global nature and low cost of entry to the Internet, innovations that are started by a teenager in one schoolhouse or bedroom can become an internationally adopted standard. These changes highlight the possibility that, with the emphasis of open source on public function testing as a means to credibility rather than credibility determined by the reputation of the code's corporate origin, youth with sufficient technical skills can enter the infrastructure development arena on an equal footing to established organizations. This implies that any attempt to prepare youth for life in the digital world should incorporate some fluency in the basic technologies of the network, and the ethical guidance in how such technologies should be implemented.

Indeed, youth have now come to expect involvement in all aspects of information in the network. Schools, business, governments, and other institutions and organizations must change to accommodate this expectation. If not, youth will migrate their attention to venues that allow for conversation and debate of information, or create their own. This is perhaps the most important implication for youth from the increasing requirement for participation on the network. Youth can and will increasingly expect to shape their information domain, and define credible information on their own terms. Any system that seeks to either impose an authority view of credibility, or that seeks to change behavior, must now be done with the understanding that youth can simply bypass these attempts and create counter-structures. Furthermore, these alternative credibility structures can have a global reach and build communities of like minds across divisions of geography, race, gender, age, and other demarcations.

Thus, there appears to be a pressing need to educate youth to assess credibility in participatory ways, thereby steering the potentially negative implications of digital networks toward positive outcomes. Because youth in particular are more self-sufficient in decision making, and also more dependent on the information that others are providing to them, digital media increase the importance of credibility, as well as their ability to make credibility judgments effectively. Learning through "conversation," typified by collaboration with others, is one avenue toward reliable credibility assessments. In the end, this is perhaps the most realistic

and effective means by which to increase knowledge among youth increasingly reliant on digital media for the information that is central in their lives.

Notes

1. Mary Madden, *Internet Penetration and Impact* (Pew Internet & American Life Project, 2006), <http://www.pewinternet.org/PPF/r/182/report.display.asp> (retrieved November 15, 2006).
2. Amanda Lenhart, Mary Madden, and Paul Hitlin, *Teens and Technology: Youth Are Leading the Transition to a Fully Wired and Mobile Nation* (Pew Internet & American Life Project, 2005), http://www.pewinternet.org/pdfs/PIP_Teens_Tech_July2005web.pdf (retrieved December 15, 2006).
3. National Center for Education, Rates of Computer and Internet Use by Children in Nursery School and Students in Kindergarten through Twelfth Grade: 2003 (2005), <http://nces.ed.gov/pubsearch/pubinfo.asp?pubid=2005111> (retrieved February 26, 2007).
4. See Gunther Eysenbach, Credibility of Health Information and Digital Media: New Perspectives and Implications for Youth, this volume, for a similar argument.
5. Jules Abend, Lands' End Uses Internet to Expand Sales, Personalize Customer Service; Bobbin, 1999, FindArticles.com, http://www.findarticles.com/p/articles/mi_m3638/is_10_40/ai_55609089 (retrieved December 15, 2006).
6. See Gunther Eysenbach, Credibility of Health Information and Digital Media, this volume.
7. Dg.o, The National Science Foundation's Digital Government Research Program, 2006, <http://www.digitalgovernment.org/> (retrieved December 15, 2006).
8. Kerry Lynn Rice, A Comprehensive Look at Distance Education in the K-12 Context, *Journal of Research on Technology in Education* 38, no. 4 (2006): 425-48.
9. Florida Virtual School, <http://www.flvs.net/educators/fact.sheet.php> (retrieved December 28, 2006).
10. Amanda Lenhart, Maya Simon, and Mike Graziano, The Internet and Education: Findings of the Pew Internet & American Life Project, 2001, 2, <http://www.pewinternet.org/PPF/r/39/report.display.asp> (retrieved December 28, 2006).
11. Miriam J. Metzger, Andrew J. Flanagin, and Lara Zwarun, College Student Web Use, Perceptions of Information Credibility, and Verification Behavior, *Computers & Education* 41 (2003): 271-90; Rice, A Comprehensive Look at Distance Education in the K-12 Context.
12. Tina Miteko, Driving Value from Every Online Customer Interaction: The Power of Intent-Driven Personalization, *Customer Inter@ction Solutions* 24, no. 8 (2006): 38.
13. Matthew L. Meuter, Amy L. Ostrom, Robert I. Roundtree, and Mary Jo Bitner, Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters, *Journal of Marketing* 64 (2000): 50-64; Allard van Riel, Veronica Liljander, and Petra Jurriens, Exploring Consumer Evaluations of E-Services: A Portal Site, *International Journal of Service Industry Management* 12, no. 4 (2001): 359-77.
14. Lane Becker, Self-Service Web Applications: Less Cost, More Value, 2002, <http://www.adaptivepath.com/publications/essays/archives/000063.php> (retrieved November 15, 2006).
15. Tina Miteko, Driving Value From Every Online Customer Interaction, 38-41.
16. Federal Election Commission, Help America Vote Act of 2002: Public Law 107-252, <http://www.fec.gov/hava/hava.htm> (retrieved November 15, 2006).

17. Grants.gov, <http://www.grants.gov/> (retrieved November 25, 2006).
18. Joanne Silverstein, *Information Technology and Commerce: Attributes of Emerging Online Business* (doctoral dissertation, Syracuse University, 1997).
19. Batya Friedman, Peter H. Kahn Jr., and Alan Borning, Value Sensitive Design and Information Systems, in *Human-Computer Interaction in Management Information Systems: Foundations*, eds. Ping Zhang and Dennis Galletta (Armonk, NY, and London: Sharpe, 2006): 348–72.
20. B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* (New York: Morgan Kaufman, 2002); B. J. Fogg, Cathy Soohoo, David R. Danielson, Leslie Marable, Julianne Stanford, and Ellen R. Tauber, How Do Users Evaluate the Credibility of Web Sites?: A Study with Over 2,500 Participants, in *Proceedings of the 2003 Conference on Designing for User Experiences*, San Francisco, CA, 2003, 1–15; Miriam J. Metzger, Andrew J. Flanagin, Keren Eyal, Daisy Lemus, and Robert M. McCann, Bringing the Concept of Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment, in *Communication Yearbook 27*, ed. Pamela J. Kalbfleish (2003), 293–335; Miriam J. Metzger, Andrew J. Flanagin, and Lara Zwarun, College Student Web Use, 271–90; C. Nadine Wathen and Jacquelyn Burkell, Believe It or Not: Factors Influencing Credibility on the Web, *Journal of the American Society for Information Science and Technology* 53, no. 2 (2002): 134–44; Joseph B. Walther, Zuoming Wang, and Tracy Loh, The Effect of Top-Level Domains and Advertisements on Health Web Site Credibility, *Journal of Medical Internet Research* 6, no. 3 (2004): e24, <http://www.jmir.org/2004/3/e24/> (retrieved January 3, 2006).
21. Andrew J. Flanagin and Miriam J. Metzger, Digital Media and Youth: Unparalleled Opportunity and Unprecedented Responsibility, this volume; B. J. Fogg, Jonathan Marshall, Othman Laraki, Alex Osipovich, Chris Varma, Nicholas Fang, Jyoti Paul, Akshay Rangnekar, John Shon, Preeti Swani, and Marissa Treinen, What Makes Web Sites Credible? A Report on a Large Quantitative Study, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seattle, WA, March 31–April 4, 2001, 61–68.
22. Andrew J. Flanagin and Miriam J. Metzger, The Role of Site Features, User Attributes, and Information Verification Behaviors on the Perceived Credibility of Web-based Information, *New Media and Society* 9, no. 2 (2007): 319–42; B. J. Fogg, Jonathan Marshall, Alex Osipovich, Chris Varma, Othman Laraki, Nicholas Fang, Jyoti Paul, Akshay Rangnekar, John Shon, Preeti Swani, and Marissa Treinen, Elements That Affect Web Credibility: Early Results from a Self-Report Study, *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, The Hague, Netherlands, 2000, 287–88; Fogg et al., What Makes Web Sites Credible?
23. Melody Y. Ivory and Rodrick Megraw, Evolution of Website Design Patterns, *ACM Transactions on Information Systems* 23, no. 4 (2005): 463–97.
24. R. David Lankes, *Building and Maintaining Internet Information Services: K–12 Digital Reference Services* (Syracuse, NY: ERIC Clearinghouse on Information & Technology, 1999).
25. Metzger et al., Bringing the Concept of Credibility for the 21st Century.
26. Frances Jacobson Harris, Challenges to Teaching Credibility Assessment in Contemporary Schooling, this volume.
27. Fred W. Weingarten, Credibility, Politics, and Public Policy, this volume.
28. See, e.g., Junghoo Choo and Sourashis Roy, Impact of Search Engines on Page Popularity, in *International World Wide Web Conference: Proceedings of the 13th International Conference on World Wide Web* (New York: ACM Press, 2004), 20–29.
29. Choo and Roy, Impact of Search Engines on Page Popularity, p. 1.

30. See S. Shyam Sundar, The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility, this volume.
31. Abbe Mowshowitz and Akira Kawaguchi, The Consumer Side of Search: Bias on the Web, *Communications of the ACM* 45, no. 9 (2002): 56-60.
32. Matthew S. Eastin, Toward a Cognitive Developmental Approach to Youth Perceptions of Credibility, this volume; Flanagin and Metzger, The Role of Site Features.
33. Flanagin & Metzger, The Role of Site Features.
34. Thomas Chesney, An Empirical Examination of Wikipedia's Credibility, *First Monday* 11, no. 11 (2006), <http://www.firstmonday.org/issues/issue11.11/chesney/index.html> (retrieved January 4, 2006); Jim Giles, Internet Encyclopedias Go Head to Head, *Nature* 438 (2005): 900-1; Felix Stadler and Jesse Hirsh, Open Source Intelligence, *First Monday* 7, no. 6 (2002), <http://firstmonday.org/issues/issue7.6/stadler/index.html> (retrieved January 4, 2007); Aaron Weiss, The Power of Collective Intelligence, *netWorker* 9, no. 3 (2005): 16-23.
35. Tanni Haas, From "Public Journalism" to the "Public's Journalism"? Rhetoric and the Reality in the Discourse on Weblogs, *Journalism Studies* 6, no. 3 (2005): 387-96; Thomas J. Johnson and Barbara K. Kaye, Wag the Blog: How Reliance on Traditional Media and the Internet Influence Credibility Perceptions of Weblogs among Blog Users, *Journalism and Mass Communication Quarterly* 81, no. 3 (2004): 622-42.
36. Deborah L. McGunines, Honglei Zeng, Paulo P. da Silva, Li Ding, Dhyanesh Narayanan, and Mayukh Bhaowal, Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study, *Proceedings of the Workshop on Models of Trust for the Web, WWW2006*, May 22-26, Edinburgh, UK, 2006, http://ebiquity.umbc.edu/file_directory/papers/274.pdf (retrieved January 4, 2007); Nikolaos Th. Korfiatis, Marios M. Poulos, and George Bokos, Evaluating Authoritative Sources Using Social Networks: An Insight from Wikipedia, *Online Information Review* 30, no. 3 (2006): 252-62.
37. Susan L. Bryant, Andrea Forte, and Amy Bruckman, Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia, in *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work*, Sanibel Island, FL, 2005, 1-10.
38. See Eysenbach, Credibility of Health Information and Digital Media.
39. Rob Law, Kenith Leung, and R. James Wong, The Impact of the Internet on Travel Agencies, *International Journal of Contemporary Hospitality Management* 16, no. 2 (2004): 100-7; Clive Wynne, Pierre Berthon, Layland Pitt, Michael Ewing, and Julie Napoli, The Impact of the Internet on the Distribution Value Chain: The Case of the South African Tourism Industry, *International Marketing Review* 18, no. 4 (2001): 420-31.
40. James E. Littlefield, Yeqing Bao, and Don L. Cook, Internet Real Estate Information: Are Home Purchasers Paying Attention to It? *Journal of Consumer Marketing* 17, no. 7 (2000): 575-90; Risa Palm and Michelle Danis, The Internet and Home Purchase, *Tijdschrift voor Economische en Sociale Geografie* 93, no. 5 (2002): 537-47.
41. Steven D. Levitt, and Stephen J. Dubner, *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything* (New York: William Morrow, 2005).
42. Eysenbach, Credibility of Health Information and Digital Media.
43. Ibid.
44. Carrie Lowe, Yale University's PACE Research Center and the Big6™, 2005, <http://big6.com/showarticle.php?id=459> (retrieved November 15, 2006).
45. See Eysenbach, this volume, for a similar argument.

46. For a greatly expanded discussion of attributes of authority and countervailing factors, see Richard Thomas De George, *The Nature and Limits of Authority* (Lawrence: University Press of Kansas, 1985).
47. Andrew J. Flanagin, Craig Flanagin, and Jon Flanagin, Technical Code and the Social Construction of the Internet, unpublished manuscript.
48. Jeffrey Zeldman, *Designing With Web Standards* (Berkeley, CA: New Riders Press, 2003).
49. Richard Wray and Dan Milmo, AOL Tries to Survive by Going Free, 2006, <http://business.guardian.co.uk/story/0,1835908,00.html> (retrieved January 2, 2007).
50. Eric S. Raymond, *Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary* (Sebastopol, CA: O'Reilly Media, 2001).
51. Galen Fott, Adobe Photoshop CS3 Beta, 2006, <http://www.pcmag.com/article2/0,1895,2074376,00.asp> (retrieved January 6, 2007).
52. See Harris, Challenges to Teaching Credibility Assessment.
53. For an example, see PEACEFIRE, 2006, <http://www.peacefire.org/> (retrieved December 29, 2006).
54. Market Wire, U.S. Teens Graduate from Choosing IM Buddy Icons to Creating Elaborate Social Networking Profiles, According to Nielsen//NetRatings, 2006, <http://www.marketwire.com/mw/release.html.b1?release.id=171656> (retrieved December 28, 2006).
55. Amanda Lenhart, Lee Rainie, and Oliver Lewis, Teenage Life Online: The Rise of the Instant-Message Generation and the Internet's Impact on Friendships and Family Relationships, 2001, http://www.pewinternet.org/PPF/r/36/report_display.asp (retrieved December 15, 2006).
56. Maya Simon, Mike Graziano, and Amanda Lenhart, The Internet and Education: Findings of the Pew Internet and American Life Project. 2001, p. 6. http://www.pewinternet.org/PPF/r/39/report_display.asp.
57. Gordon Pask, *Conversation Theory: Applications in Education and Epistemology* (New York: Elsevier, 1976).
58. Andrea Hemetsberger and Christian Reinhardt, Learning and Knowledge-Building in Open-source Communities: A Social-Experiential Approach, *Management Learning* 37, no. 2 (2006): 187–214.
59. See Harris, Challenges to Teaching Credibility Assessment.
60. Peter S. Jenkins, Leafleting and Picketing on the "Cydewalk" – Four Models of the Role of the Internet in Labour Disputes, *UCLA Journal of Law and Technology* 7, no. 1 (2003).
61. See Eastin, Toward a Cognitive Developmental Approach to Youth Perceptions of Credibility.
62. Wikipedia, Astroturfing, 2006, <http://en.wikipedia.org/wiki/Astroturfing> (retrieved December 29, 2006).
63. <http://digg.com/>.
64. Chris Davies, Geoff Hayward, and Linariah Lukman, *Digital Technologies: A Review of Research and Projects* (2006): 14–19, http://www.futurelab.org.uk/download/pdfs/research/lit-reviews/futurelab-review_13.pdf (retrieved December 29, 2006).
65. Sundar, The MAIN Model.
66. Soo Young Rieh and Brian Hilligoss, College Students' Credibility Judgments in the Information-Seeking Process, this volume.

Trusting the Internet

121

67. Brett D. Jones and Robert J. Egle, Looking through Different Lenses: Teachers' and Administrators' Views of Accountability, *Phi Delta Kappan* 87, no. 10 (2006): 767-71; Harris, Challenges to Teaching Credibility Assessment.
68. Harris, Challenges to Teaching Credibility Assessment.
69. Weingarten, Credibility, Politics, and Public Policy.
70. Harris, Challenges to Teaching Credibility Assessment.
71. Ibid.
72. Miriam Metzger, Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research, *Journal of the American Society for Information Science and Technology* 58, no. 10 (2007).
73. Harris, Challenges to Teaching Credibility Assessment.
74. See Weingarten, Credibility, Politics, and Public Policy.

AUTHOR(S):

Zollers, Alla
Moore, Tony
Nguyen, Lilly
Russell, Terrell
Stutzman, Fred

ORGANIZATION(S):

University of California, Los Angeles, United States of America
Drexel University, United States of America
University of North Carolina, Chapel Hill, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Overview

Social tagging has emerged as one of the most popular social software tools available online. Originating from Delicio.us, social tagging capabilities can now be found on a number of major music, news, video, and commercial websites, as well as on social network sites and enterprise systems. Although social tagging allows individuals to organize content utilizing user-generated vocabulary, the power of social tagging stems from the ability to view and share resources with other users of the system. Through the sharing of tags and resources, social tagging systems facilitate network connections and perhaps even the creation of communities.

In this panel, an exciting group of young researchers will present their ongoing work on social tagging. This panel will present a variety of perspectives on social tagging ranging from qualitative ethnographic work to quantitative visualizations. Additionally, the panel will cover topics such as: the definition of a tag, the role that tags play in social network sites, as well as tags in corporate and organizational settings.

The research and the varying methods presented in this panel will present viewers with an exciting array of perspectives on social tagging. Additionally, in order to further engage the audience, the panelists will also participate in a point-counterpoint discussion with the participants which will help illuminate both the advantages and disadvantages of social tagging, as well as further highlight the multiple perspectives and approaches available for continuing social tagging research.

What is a Tag?

Alla Zollers

The social software revolution is about harnessing the wisdom of crowds by aggregating the small contributions of millions of people. Social tagging is one aspect of this revolution because it allows individuals to define and classify the world in their own terms. Tags, which form the most fundamental part of a social tagging system, have mostly been understood in terms of metadata. However, this talk will demonstrate that for social tagging to be considered part of the social software movement, the tags themselves need to encompass a much broader and social definition. The definition of a tag can be extended include constructs such as: tags as informal annotations, tags as network links, and tags as representations of self, community, and values. All of these facets of tags provide the social, communal, and collaborative underpinnings of social tagging.

Opportunities and Applications of Corporate Social Tagging

Tony Moore

The organization of digital information is subject to social and organization factors such as context, come into play with all design engagements. As such, it is rational to conclude that since design [information organization and access] exist in a social context the results can have unintended purposes. There are various approaches and techniques for integrating user input into the management and organization of digital information. With each approach comes a mixture of advantages and disadvantages. This research briefly discusses popular and varying approaches to user engagement in digital information organization. This underlying theoretical approach is based on an assumption that information and knowledge grow through dialogue and communal understanding of issues. I conclude with two varying example of tagging behavior in a consulting firm. It is the goal of this work to document existing themes in tagging behavior and support further research in this area.

Classification in Context: An Ethnography of Tagging Practices

Lilly Nguyen

This paper seeks to contextualize the current research in social classification systems through an ethnographic approach to understanding the ways in which individuals tag, classify, and the meaning of these practices. This paper argues for an understanding of online social classification systems as social practice, in contrast to previous frameworks for classification within a cognitive psychology frame that have understood classification as perception. As such, following sociological and science, technology, and society (STS), this paper explores the ways in which social classification systems represent the underlying schemes of social organization in networked digital life.

Tag Decay

Terrell Russell

Terrell Russell is a PhD Student in SILS at UNC-Chapel Hill. He is working on a social tagging method of expertise discovery in organizations. One of the most interesting aspects of this discovery is working through the problem related to out-of-date information. Studying Tag Decay may provide some answers to questions such as: Are there types of knowledge that do not go stale? Can the stability of the set of descriptors for an object be measured?

Social Annotation in Social Network Sites

Fred Stutzman

Social network sites, used by millions of individuals, support a wide range of uses across diverse populations. As these networks are largely ego-centric, one of the primary user behaviors is self-identification and identity construction. At a primary level, identity is constructed through self-description of interests, hobbies and personal preferences; these elements are realized in the interface through browsable collections of hyperlinks. Due to the representative nature of these hyperlinks, we can think of these descriptive elements as self-tags, a form of social annotation. This behavior raises many interesting questions. What sorts of self-tags are most popular? What are significant or unusual examples of self-tags? And what motivations do users have for tagging themselves in a social network site? Drawing on a large dataset collected from Facebook.com, I will explore the practice of social annotation in social network sites.

WILDCARD SUBMISSION 6:

Author Meets Critics: Sandra Braman

AUTHOR(S):

Blanchette, Jean-François
Braman, Sandra
Jackson, Steve
Lievrouw, Leah
Mueller, Milton

ORGANIZATION(S):

University of California, Los Angeles, United States of America
University of Wisconsin-Milwaukee; United States of America
University of Michigan, United States of America
Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The publication of Sandra Braman's *Change of State* signals the importance of the topic of information policy to the field of information studies. It also illustrates both the power and necessity of the kind of inter-disciplinary analysis that characterizes the field. This panel will provide an opportunity for both structured debate and lively discussion around *Change of State* and the important arguments it contains. Perhaps the most important one that it makes is that "trends in information policy both manifest and trigger change in the nature of governance itself." That is, Braman points to how information flows have radically altered the nature of the traditional nation-state and its ability to exercise power (or fails to, as the case may be).

Because of i-Schools' strong connections with the professional world of information institutions, it is particularly important for graduates to understand the far-reaching implications of this argument. On a practical level, graduates need to become conversant with both the varied manifestations of information policy, and the mechanisms by which such policy is enacted. *Change of State* provides a highly useful synthesis of all the major debates in the field of

information policy (including identity, immigration, innovation policy, etc.), organizing them into a comprehensive analytical framework, along with extensive bibliographic essays.

The format of this session will involve each speaker presenting a 15 minute paper which will discuss Change of State from their own disciplinary perspective. Prof. Braman will then respond to the papers. The presentations will be followed by a question and answer period with the audience. The goal is to offer both structured debate around key arguments presented in the book, while offering an opportunity for the audience to interact with the speakers. The speakers are all leaders in the field, well-informed of Dr. Braman's work, and chosen with the goal of maximizing disciplinary perspectives (Lievrouw: communication policy; Mueller: political science; Jackson: science and technology studies).

Participants:

- Prof. Sanda Braman, Dept. of Communication, University of Wisconsin-Milwaukee, <http://www.uwm.edu/~braman/>
- Prof. Steve Jackson, School of Information, University of Michigan, Ann-Arbor, <http://www.si.umich.edu/people/faculty-detail.htm?sid=366>
- Prof. Leah Lievrouw, Department of Information Studies, UCLA, <http://polaris.gseis.ucla.edu/llievrou/LeahHome.html>
- Prof. Milton Mueller, School of Information Studies, Syracuse University, <http://istweb.syr.edu/~mueller/>

"Change of State" page at MIT press: <http://mitpress.mit.edu/catalog/item/default.asp?tttype=2&tid=10925>

Because of constraints on speakers' schedules, this session must imperatively be scheduled for Thursday, February 28.

WILDCARD SUBMISSION 7:

CHI, ICIS, SIGIR and Numerous Other Outlets That Won't Accept My Work: The Perils of Publishing Multidisciplinary Research?

AUTHOR(S):

Jansen, Bernard

ORGANIZATION(S):

Pennsylvania States University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The motivation for this session comes directly from two series of comments made at the 2007 iConference. The first was uttered at a pre-conference workshop. In a discussion on where to publish, the recommendation was made to aim for the outlet First Monday. Now First Monday is a fine online journal, but it is certainly not on any institution's A-list that I know of. If First Monday is your lead journal, what's your fall back?

The second comment was made at a session on digital library research, where the panelists were asked what they considered their major publication outlets. The overwhelming answer was the SIGIR Conference. Now SIGIR is fine conference, but when one thinks of openness to new ideas, SIGIR is not the first conference that jumps to mind.

These comments, and many others made in informal discussions, raise the concerns of disseminating multidisciplinary research results.

This interactive session will address the subject of publishing multidisciplinary research. By multidisciplinary, we also include interdisciplinary, transdisciplinarity, crossdisciplinary, cross cutting, multidisciplinary and various other terms that academics and others use to describe:

"the act of drawing from and integrating two or more academic disciplines, professions, technologies, departments, their methods and insights, in the pursuit of a common goal. Interdisciplinary approaches typically focus on problems felt by the investigators to be too complex or wide-ranging to be dealt with using the knowledge and methodology of a single discipline" (Wikipedia, <http://en.wikipedia.org/wiki/Interdisciplinary>).

The question of the challenges of multi-disciplinary research is not new. There has been much written about multidisciplinary programs and education (see Chubin, 1976 or Newell, 2001 as examples). Most of this work has focused on designing of programs. There has been some articulation of the barriers to engaging in multidisciplinary Jansen

research, with the concern that multidisciplinary research can be seen as lacking in rigor (from the viewpoint of established disciplines). It has been noted that the tendency is for interdisciplinary research to become disciplines

themselves. One could make the case that computer human interaction has followed such an approach. We see this now with the iSchools. The iConference series has moved from iSchools the community to defining iResearch, iMethods, and the iSchool space.

However, there has been much less formal discussion concerning the publication of multidisciplinary research. The subject is nuanced. For example, it is difficult to find a journal or conference that does not advertise itself as 'seeking research from a broad range of perspectives on' [insert domain here]. However, antidotal evidence suggests that many times there are severe qualifiers to these welcoming solicitation statements. The impression expressed by some multidisciplinary researchers is that journal and conferences are open to new perspectives ... as long as such research uses the field's assumptions, methods, and literature.

As such, many academics claim that barriers do exist in publishing multidisciplinary research.

If this criticism is justified, what can be done about improving the publication opportunities for multidisciplinary researchers? It would seem that this is a critical issue for the iSchools for future recruitment and retainment.

However, is this criticism justified? Or, is it just sour grapes for research that is of little impact or low quality? Is the solution just to 'published good research'? If this is the case, how do multidisciplinary institutions go about correctly a common misperception that multidisciplinary research is challenging?

These are the motivators for this session. We aim to examine this subject objectively, critiquing the major viewpoints, and producing some avenues for possible future courses of action. As such, we do not aim for this session to be a 'whine-fest'. Rather, the session is structured to constructively engage and produce some directions to either deal with improving the publication process for the iSchools or putting the matter (i.e., misperception) to bed.

Issues

This interactive panel will address three issues in the 1 ½ hour session, which are:

1. Are there unique barriers to publishing multidisciplinary research?
 - a. If there are, what are they?
 - b. If there are not, why is there the misperception that it is so commonplace?
2. Depending on the answer to (1)
 - a. What can be done to over come the barriers of publishing multidisciplinary research? (and who needs to do it?)
 - b. What can be done to correct the misperception concerning barriers to publishing multidisciplinary research? (and who needs to do it?)
3. Is there a need for iResearch, iMethods, and an iSchool space?

2 of 5 Jansen

Depending on the answer to (3)

- a. If there is a need, what are the iResearch, iMethods, and iSchool space?
- b. If there is not, what do the iSchools bring to the table in terms multidisciplinary work that has not existed previously?

Contributors

- Facilitator: Dr. Jim Jansen (Penn State)
- Panelists: We have intentionally recruited panelists who hold diverse opinions concerning these issues, which should make for an informative session.
- Panelists and positions are:

Andrew Dillon (quite whining and just publish good research)

Dean, School of Information

Professor of Information, Psychology, and Information, Risk & Operations Management

School of Information

SZB 564

1 University Station D7000

University of Texas, Austin, TX 78712-1276

Email: adillon@ischool.utexas.edu

Howard Rosenbaum (publishing multidisciplinary research is really hard)

Associate Professor of Information Science

School of Library and Information Science

1320 East 10th Street

Indiana University

Bloomington, IN 47405

Email: hrosenba@indiana.edu

Mark S. Ackerman (middle of the road)

Associate Professor

Department of Electrical Engineering and Computer Science and School of Information

University of Michigan

Ann Arbor, Michigan 48109

Email: ackerm@umich.edu

- Participants: The format of the session will be designed to actively engage the session attendees.

Format

1. Opening Remarks: The facilitator will present opening remarks (2 ½ minutes)
2. Session Survey: We will begin the session with a self-reflection on the three issues listed, with each participant and panelist individually addressing the three issues, writing down responses (10 minutes).

3 of 5 Jansen

These written responses will be collected and quickly open coded while the panelists are giving short opening remarks.

3. Panelist Opening Remarks: Each panelist will give opening remarks concerning the three issues mentioned above (10 minutes each, 30 minutes).
4. Discussion: We will present the results from the participant and panelist surveys, and then use these results as a basis for a question and answer period among participants and panelists (30 minutes).
5. Outcomes and Final Statements: Based on various positions, discussions, and questions, each panelist will present their prescription on what a course of action should be for the iSchools (5 minutes each, 15 minutes).
6. Concluding Remarks: The facilitator will present concluding remarks (2 ½ minutes)

Outcomes

The session will have administrative support to accomplish three things:

1. Aggregate, analyze, and provide the responses to the self-reflection exercise from each session participant to the Dean of each of the iSchools and others as requested.
2. Summarize the discussion period comments with a focus on concerns, successes, and barriers.
3. Summarize the recommendations on courses of action and provide to the Dean of each of the iSchools.

Needed Support from the iConference

- Student volunteers to distribute, collect, and transcribe the surveys. Student support also needed for note taking during the session.
- A computer or two for the students to code the data.
- A bunch of pencils.

Review Criteria

The topic addressed in this proposal is a critical component of the conference theme of multi- (or inter- or cross-) disciplinarily in all areas, including participants, literatures used, and methods. The issue of whether or not the iSchools offer anything unique is something that must be addressed. Additionally, untenured faculty have to come to grips with the true nature of being in an iSchool – whether multidisciplinary research is a boon, a bust, or a non-issue for tenure.

This session will exclusively focus on publishing multidisciplinary research. Publications are in fact the tangible result of research, multidisciplinary or otherwise. In order for a multidisciplinary collection of faculty to be ultimately successful, they must understand the underlying process of publishing, be that in disciplinary outlets or in multidisciplinary venues. For tenure track faculty to be ultimately successful, they must find thriving processes for both publishing and for the related areas of grants.

4 of 5 Jansen

This is an area of interest and concern for current faculty, Deans of the iSchools, for recruiting, and graduate students. As such, the session will be domain spanning.

WILDCARD SUBMISSION 8:

Info-Scavengers

AUTHOR(S):

Rosner, Daniela Karin
 Lim, Kevin Mateo
 Hesse, Hannes

ORGANIZATION(S):

University of California, Berkeley, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Scavenger hunts are emergent collections, constructed out of small bits of information gathered by the players. As information scholars, we engage in scavenger hunts every day, whether researching ideas to formulate a paper, looking for the right regex pattern to parse a chunk of data, or integrating information into the correct ontology for a set of documents. We hope to illuminate this process in our everyday lives as students by revisiting the scavenger hunt for the world of us, the informavours [3].

We propose a scavenger hunt that takes place in and around the UCLA library. We intend to base some of the structure of the event on Jane McGonigal's work on Big Games [2]. Participants will engage the physical environment of UCLA campus in the access of information. Groups of 4 members will be given a set of 15 clues about different types of information that exist in a multiple spatiotemporal locations. In order to unlock their clues, groups will forage [4] and retrieve information using social and theoretical processes studied in the field. Information will include literary objects in the library, files "buried" in the web, and actual physical objects around the UCLA campus. In following with the berrypicking model of information retrieval [1], some clues will point to other clues, helping participants refine their search. Such clues interact with each other and need to be decoded in a certain order; however, all 15 clues will be given out at the same time. Groups will have 1 hour to gather all 15 clues, assemble them into the right order and unlock a final clue to win the game.

As a technical requirement, teams will need access to UCLA's wireless network so they have access to the world wide web, where some of the clues will be located. Each team should have at least one laptop between them.

[1] Bates, M. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. <http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>

[2] McGonigal, J. Supergaming: Ubiquitous Play and Performance for Massively Scaled Community. *Modern Drama* 48:3, 2005.

[3] Miller, G. A. (1984). Informavours. In F. Machlup & U. Mansfield (Eds.) *The study of information: Interdisciplinary messages*. New York: Wiley.

[4] Pirolli, P. and Card, S. (1995). Information Foraging in Information Access Environments, *Proc. CHI'95*, 11 S-25

WILDCARD SUBMISSION 9:

iSchool Proposal for Themed Wildcard Session on New Information Systems Methods

AUTHOR(S):

Star, Susan Leigh
 Jirotko, Marina
 Whalen, Jack
 Calvert, Scout

ORGANIZATION(S):

Santa Clara University, United States of America
 Oxford University, United Kingdom
 Xerox PARC, United States of America
 University of California, Santa Cruz, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

'New Information Systems' is an emerging field composed of social studies of science (STS), information sciences (IS), workplace studies and technological design, and new media forms such as cyberinfrastructure or eResearch. Within this area we are exploring the connections and inter-relationships between empirical studies of information at knowledge creation and use, and methods from more traditional IS, social networks, grounded theory and ethnomethodology. The collective creation of a theoretically driven cluster at this juncture would tie us together in a convergence that would link our scholarship and enable students to access this strong and existing - yet invisible - college. We propose a "wildcard" session here that makes a space for people to speak about their methods, assess their viability for helping to build our emerging community, and hopefully to explore the "behind the scenes" actions associated with practicing any methods.

Such an event is most timely. At the recent meetings of the Society for the Social Studies of Science (4S), an entire day was devoted to the emerging intersections of STS and IS. In addition it should be noted that the same book, *Memory Practices in the Sciences* (G. Bowker, MIT Press, 2007), won the best book award at both ASIST and 4S. This might be seen as a harbinger for the deeply theoretical and methodological work that is to take place, if the intersection is to be a robust one.

Our research directions will be focused on studies of infrastructure, ethical actions that are inscribed into IS, and theoretical studies of questions such as 'what is useful information?' We need to unpack the contextual nature of knowledge creation and use. As well, we need to understand the ways in which it is entangled with obligations from different domains and communities of practice such as privacy, consent, anonymity, confidentiality, ownership and a whole host of organizational and professional matters. New media studies point to an intense overlapping and interrelationship of fields and disciplines.

Methods should come from a combination of (1) sensitivity to the historical moment (e.g., multiculturalism, extreme changes in the meaning of 'global'); (2) an assemblage of tools that are ready to hand, theoretically driven, are pleasant and effective to use; and (3) embody an ethical commitment to the values and meanings of those who are being studied (emic), within a way to explore the conventions, standards and infrastructures that both constrain and enable their experiences (etic).

The papers here aim to show a range of approaches from the current STS, IS and Workplace Studies emergence that speak to the criteria detailed above. Each participant in the experiential forum will bring an example of their research, and as honestly as possible, assess its methodological strengths and weaknesses. The assessment will be relative to strengthening the development of the iSchool community, to the intersections noted above, and to the welfare of respondents.

Susan Leigh Star (Santa Clara University)

Assessing grounded theory and boundary objects. Grounded theory is the most popularly cited analytic approach in qualitative methods. Boundary objects is a notion, found in the work of Star and co-authors, that examines assemblages of humans and things, and how things that exist in more than one community of practice are used in performing cooperative work. The discussion here attempts to combine the information/knowledge needs and cooperative work aspects of boundary objects, with the theory-generating power of grounded theory. Leigh will explore what such a theory may look like, and what boundary objects may be used in this combined approach. She will detail how to study both intersecting and shared communities, and how to develop theory by constant comparison, lateral thinking and the poetics of infrastructure.

Jack Whalen (Xerox PARC) and Marina Jirotko (Oxford University)

This section will discuss the sensitivity of an ethnomethodological orientation for revealing the practical accomplishment of knowledge creation and use – in its moment to moment unravelling of the complexities of these knowledge practices. Jack and Marina look at how members of different work communities -- from service technicians to software engineers to vehicle manufacturing (hardware) engineers to 9-1-1 dispatchers to radiologists and financial traders-- learn what they know to do what they do, in very practical terms, and the role of this practical knowledge in organizational life.

In this regard, it seems likely that much of an organization's most valuable knowledge, its 'intellectual capital', is not centered in official document repositories and databases ? scientific formulae, "hard" research data, computer code, codified procedures, financial figures, customer records, and the like ? but rather is in the largely undocumented ideas, insights, and know-how of its members.

We have learned, through careful studies in the field, at the work site, that this quite tacit knowledge is deeply rooted in the experiences of individuals and the culture of their work communities. It commonly originates as practical solutions -- through everyday inventions and discoveries -- to the exigencies of work, and in this way serves as the critical resource for ordinary work practice. In fact, much of this knowledge often remains embodied in the practice.

Crucial steps in a new practice and fresh solutions to recurrent conundrum are commonly shared through conversations and stories among small circles of colleagues and work groups, with members filling in the blanks from their own experience. These instructions and stories are naturally expressed in the language -- the local vernacular -- of those groups and communities. But it is equally recognized that the challenge for organizations is to then somehow convert this valuable but mainly local knowledge into forms that other members of the organization can understand and, perhaps most important, act upon together.

Jack has been concentrating on how technology, in concert with changes in work processes or structures, can enable this 'conversion', this concerting and coordinating of community action. Marina's has been concerned with bringing a richer comprehension of socially organised work practice into the process of engineering technological systems that focus on supporting everyday work and interaction.

In this session, drawing upon specific empirical examples, they will describe some of these experiences along with the strengths and challenges of adopting this approach in organisational settings.

Scout Calvert (UC Santa Cruz)

My work is about the enduring qualities of the book and how its resistance to commodification enables this thing we call a public library. Even as new media forms gain prominence, they must nevertheless reckon with and be measured against the book form. I am interested in developing library-appropriate research methods for thinking critically about old and new technologies, and the way these technologies allow or constrain decisions on the part of librarians. Many of the decisions librarians make are in a fraught ethical domain: privacy and freedom of information are obvious examples. However, we naturalize the way technological infrastructures make these decisions for us. I would like to collaborate in producing methods and an ethic that allows librarians to mandate characteristics in developing technologies which both preserve librarians' prerogatives to make ethical judgment calls, and make the fact that those judgment calls are necessary visible. Specifically, this is a search for a method for integrating appropriate technologies into library work that will structurally preserve the possibility of a public library, even as the book mutates and shares the stage with other media. Ultimately, "content" only matters to the degree it is germane and hence not a generic commodity bought to fill up space (ie., it must be more than "mere" content), while "containers" continue to actually really matter and dictate the name of the game for the immensely diverse patronage of the public library.

WILDCARD SUBMISSION 10:

I-schools in Central and South Europe: Developments and Challenges of Cooperation

AUTHOR(S):

Aparac-Jelusic, Tatjana

Schloegl, Christian

Zumer, Maja

Tamamro, Anna Maria

ORGANIZATION(S):

University of Zadar, Croatia (Hrvatska)

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The countries of Central and South Europe entered the 2000s with plans to restructure their higher education systems based on Bologna recommendations and some other EU supported guidelines. The Humboldt tradition at the universities in the region, necessity to introduce new academic profiles as well as the use of ICT in higher education, lead towards restructuring of existing or designing new curricula and to cooperation between the LIS and IS departments in the region.

This panel will address the special concerns for restructuring curricula in the wider information sciences field (Library and Information Science - LIS, Information Science - IS, Records Management - RM, education for digital services etc)

Panelists will address three questions with respect to i-schools developments and trends in their respective countries:

- What are the basic concerns in developing new curricula in their respective countries?
- How do the multidisciplinary and interdisciplinary issues influence curricula design in IS?
- What are the future steps, particularly in regard to regional cooperation?

Panelists:

T. Aparac-Jelusic, Redesigning existing curricula at LIS and I-schools: Croatian perspective

Prof. Dr. Aparac-Jelusic, chairs the IS program at Osijek and LIS program at Zadar,

Croatia, and also directs the annual conference in Dubrovnik, Libraries in the Digital

Age. She will set i-schools creation and development in Croatia in the context of early attempts in 1960s and initiatives for redesign of curricula at the beginning of 2000s. She will explore how the new programs reflect the need to educate students in the field of digitization (especially in relation to heritage material and repositories of educational content). She will give special attention to the cooperation in education and research in the region.

Christian Schloegl, IS Schools in German-language countries.

Dr Schloegl is Assistant Professor at the Institute of Information Science, Karl Franzens University, Graz, Austria

Dr Schloegl will give an overview to I-Schools in German-language countries stressing specific features of information science in German-language countries. Since the I-Schools Conferences deals also with multidisciplinary and interdisciplinary issues, Dr Schloegl would introduce the Graz approach on information science and discuss the opportunities but also the threats related to a cross-disciplinary approach.

Maja Žumer, New Curriculum in Slovenia: Too much change or not enough.

Maja Žumer is the Head of Department in Ljubljana and member of several IFLA working groups.

She will describe the characteristics of the previous program and focus on the planning of the new curriculum, the discussions and dilemmas. She will discuss the need for constant evaluation in the framework of European guidelines and recommendations.

Anna Maria Tammaro

She is the Head of the LIS Chair at the Department of Arts and Heritage, involved in new curricula design and collaborative European educational projects. Ms Tammaro will initiate debate about the reform application in LIS, Archival and IS schools in the Mediterranean Area, between a traditional approach and an innovative one, based on harmonization of courses and recognition of qualifications.

WILDCARD SUBMISSION 11:

Bridging Discourses: Exploring the Relationship between Information Technologies and International Development

AUTHOR(S):

Srinivasan, Ramesh

Jackson, Steven

Parikh, Tapan

Wallack, Jessica

ORGANIZATION(S):

University of California, Los Angeles, United States of America

University of Michigan, United States of America

University of California, Berkeley, United States of America

University of California, San Diego, United States of America

Center for Development Finance, India

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

The panel will bring together scholars of information with training in computer science, economics, cultural studies, and the interdisciplinary social sciences to explicitly probe into the relationships held between information and

technology initiatives and international development studies. The scholars presenting are leading voices conducting global applied research that concurrently consider social, cultural, economic, and technological factors. We believe that this interdisciplinary set of short presentations can stimulate an important dialogue related to the emergent relationships held between information and the over 70% of the world that still lacks routinized access to digital technologies and the internet. The presenters shall be:

1. Ramesh Srinivasan, Assistant Professor, UCLA, Department of Information Studies: Dr. Srinivasan shall present his research based in rural Andhra Pradesh, India (and conducted in other global locales to explore how information technologies can be used as reflective devices in stimulating rural decision-making and development conceptualization. He shall present data that show how technologies may be culturally localized and serve as stimulants by which rural peoples can build capacities, strategically make decisions around different development priorities, and build new relationships with governmental, institutional, and commercial forces that are emerging in the developing world. With this perspective he will problematize the often conflicting goals in IT-development objectives between local sustainability and economies of scale.

2. Tapan Parikh, Assistant Professor, UC Berkeley I-School: As a computer scientist who has designed a number of mobile technologies in the developing world, Parikh has found that designing and evaluating technologies for the developing world presents unique contextual challenges – including limited time and resources in the field, lack of infrastructure and cultural differences that lead to misunderstandings and inefficiency. In this panel, I would like to discuss my recommendations for establishing relationships with local partners.

Establishing Rapport

Establishing rapport is the first essential step for the success of a design project. Local partners understand the context and can achieve results more effectively than outsiders. Establishing rapport requires the design team to show they are committed and capable; and providing partners the opportunity to do the same.

“Show you are tough.” We traveled regularly on the chicken bus from Guatemala City to Barillas and back (over 14 hours each way). We have walked in the mud for hours, slept on wooden boards with farmers in insect-infested sub-tropical forests. This establishes the precedent that we are working on equal terms.

“Show you are capable.” We arrive promptly at the office every day, and show that we are committed. We are careful to ask relevant questions, and stay focused on the task. We refer to experience in other countries, which arouses confidence and interest.

“Clearly establish roles.” It is important to be realistic about what you can provide, what you expect local partners to provide, and what is the potential for benefit. Being open and honest is the best way to avoid unrealized expectations.

Overcoming Evaluation Challenges

It is difficult to evaluate computing systems in the developing world. Obtaining useful and statistically relevant data requires patience and planning.

“At first, be patient and flexible.” Due to rigorous travel, the design process should coincide with scheduled activities wherever possible. Most of our initial studies are organized with field visits, which leaves limited time and attention for design work.

“Conduct experiments that are clearly relevant for participants.” The best way for participants to sincerely participate is to convince them that their efforts will have long-term benefits. This focuses the research team on those questions that have the most immediate and direct impact.

“Strive for a balance between control and enthusiasm.” It is difficult to maintain a sterile testing environment - there are dozens of people around, and users get distracted. While this hinders the design process, natural interest and excitement also helps build rapport and user involvement.

“Take compliments with a grain of salt.” People from many cultures are polite and will not plainly give a negative answer. Some participants are also keenly interested in acquiring new technology (for example, fancy mobile phones).

“Use accumulated social capital to achieve better testing conditions.” In India, once we established a long-term working relationship, we were able to conduct experiments under controlled conditions [1]. We are hoping for the same opportunities with future partners.

Discussion

With success comes responsibility. We have been afforded an opportunity to work with local communities in the expectation of mutual benefit. Carrying out projects to a sustainable hand-off point is the next challenge. Without local adoption and ownership of systems, the end goal is still some distance away.

3. Jessica Wallack, Assistant Professor, UC San Diego - Department of International Relations and Pacific Studies (IRPS), and Center for Development Finance (Chennai, India): Revealing Freedoms: ICTs, Market Infrastructures, and User Communities

The quest to expand human "freedoms" has become a central focus of development policy and research over the past decade. Amartya Sen's 1999 *Development as Freedom* catalyzed an already emerging consensus that development was as much as about providing choices and opportunities as it was about increasing incomes or wealth.

Many development-oriented proposals focus on creating choices and opportunities, but ICTs hold great potential for contributing to development by revealing freedoms. Ebay, Craigslist, Guru.com, and other web-based trading systems have created entirely new markets for goods and services – and new opportunities for buyers and sellers – simply by creating a platform for existing demand and supply to be articulated more efficiently across social networks, place, and time.

The ICTs themselves cannot reveal choices and opportunities, however: the systems must engage users to be effective. How? How can designers get people to use the infrastructure by making sincere offers or placing honest orders? How can we encourage them trust the offers/orders that are coming from the system itself? How can we catalogue the choices and opportunities so that users can easily find the relevant ones? The search engines and trust-building features like Paypal's escrow accounts, Ebay's user ratings, and Craigslist's user "flagging" of inappropriate postings were critical design features.

This paper will analyze analogous challenges and solutions for ICT market infrastructures designed to reveal choices and opportunities for low-income, remote, rural, emerging technology users. The challenges are in part technical – "technology" is more likely to be a cell phone with limited data capabilities and a basic user interface than a computer with high speed internet – but also social. Technology is a more recent introduction among these communities. Existing transactions may have additional meaning than just the economic exchange, creating resistance to any disintermediation of the face to face exchanges that goes along with increasing the number of transactions.

It will discuss preliminary lessons on solutions drawn from a survey of ICT market platforms in developing countries, as well as the author's experience with an ongoing effort to develop and deploy a cell phone based market information structure in rural Tamil Nadu, India. The paper will also outline that project's methodology for a more systematic assessment of what works.

4. Steven Jackson, Assistant Professor, University of Michigan, I-School: This paper will present early findings from the African Knowledge Infrastructures project, which seeks to broaden and reconceptualize IT for development practice in an African context. The paper will explore failings and pathologies widespread within IT for development efforts to date, and suggest ways in which new thinking and practice around the categories of knowledge and infrastructure might call into vision new styles and possibilities for developmental practice.

WILDCARD SUBMISSION 12:

Okay, Facebook me: Exploring behavior, motivations and uses in Social Network Sites

AUTHOR(S):

Stutzman, Frederic
Boyd, Danah
Marwick, Alice
Lampe, Cliff

ORGANIZATION(S):

University of North Carolina, Chapel Hill, United States of America
University of California, Berkeley; United States of America
New York University, United States of America
Michigan State University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Format: In the information sciences, there is significant interest in the study of social network websites. Sites such as Myspace and Facebook have drawn hundreds of millions of users, shaping and contextualizing new forms of social information transfer.

Furthermore, the impact of social networking sites on the college campuses we inhabit is substantial; students use social network sites to connect, engage with fellow students, and to form new learning opportunities.

In this wildcard session, four (with a potential fifth) researchers exploring social network sites will come together to debate and examine the future of social network scholarship. The panelists bring a significant breadth of experience, varying research questions, and a broad scope of methodological approaches. Rather than simply presenting research and fielding questions, the panelists will engage in a debate exploring some critical and exploratory aspects of social network sites. What do social network site behaviors represent? How are we constructing new publics in social network sites? What is the future of social networking? With this high interactivity format, this session will provide significant insight into social network sites.

Scheduled to participate in the session are danah boyd, Ph.D. student at UC-Berkeley's School of Information, Cliff Lampe, Assistant Professor in the department of Telecommunications, Information Studies and Media at Michigan State University, Alice Marwick, PhD student at New York University in the Department of Media, Culture, and Communication, and Fred Stutzman, Ph.D. student at the University of North Carolina at Chapel Hill's School of Information and Library Science.

Biographies:

danah boyd is a Ph.D. student in the School of Information at the University of California, Berkeley and a fellow at the Annenberg Center for Communications at the University of Southern California. Her research focuses on how people negotiate a presentation of self to unknown audiences in mediated contexts; her dissertation is looking at how youth engage with networked publics like MySpace, Facebook, LiveJournal, Xanga and YouTube. She is interested in how the architectural differences between unmediated and mediated publics affect sociality, identity and culture. danah has an A.B. in computer science from Brown and a M.S. in sociable media from MIT Media Lab.

Cliff Lampe is an Assistant Professor in the department of Telecommunications, Information Studies and Media at Michigan State University. His research interests include the use of rating systems in online conversations, how online interactions are translated into offline activity, and online user misbehavior. Besides working with Slashdot, Cliff has studied online communities such as Everything2, Newstrust, and Facebook. Cliff received his doctorate from the University of Michigan's School of Information. In his research, Cliff explores how college aged people using Facebook have experienced benefits from their use of the site, maintaining relationships that they might otherwise have abandoned. These relationships, often tenuous and ephemeral, provide social capital benefits that can be enacted in the future. Consequently, relationships maintained through Facebook stretch the definitions of "friends" to a wide range of nuanced relationships. Also, as users older than the initial academic participants join the site, what experiences are they having? Are they seeing the same social capital benefits as their college-aged counterparts, or having other experiences altogether?

Alice Marwick is a PhD student at New York University in the Department of Media, Culture, and Communication. She studies social media from a critical perspective, including the political economy of new media, online identity, and feminist media theory. Her primary analytical lens is to view internet applications as commercial structures: how does a site's business model affect the types of user actions that are encouraged or discouraged? Social networking sites typically circumscribe user actions, such as page customization, based on what is economically beneficial for the parent company. Values often associated with social networking sites, such as social interaction, privacy, participation, and creativity, must therefore be understood within a commercial context.

Recent work interrogates claims of YouTube as an inherently democratic space through an examination of community norms and Google's corporate practices. Alice's dissertation will examine the creation and maintenance of status hierarchies within social media through ethnographic work on the so-called "Web 2.0" community of the San Francisco Bay Area. She has worked in the technology industry since 1995 as a researcher, writer, and project manager.

Fred Stutzman is a Ph.D. student at the University of North Carolina at Chapel Hill's School of Information and Library Science. His research interests include social computing, online social networks and digital identity. Since 2005, he has explored user behavior in online social network sites, primarily Facebook.com. He has conducted a number of studies exploring topics such as adoption, privacy and the network structure in online social network sites. Currently, he is exploring data from a 16-week study of college freshman adoption of Facebook. In this session, Fred plans to share information about user motivations and behaviors in online social network sites, as well as describing research methods and challenges of studying online social networks. In addition to his graduate work, Stutzman is the co-founder of ClaimID.com, the identity-management website.

WILDCARD SUBMISSION 13:

"Let's get Wild: Building a National Research and Service Agenda for Community Technologies and Networking"

AUTHOR(S):

Crandall, Michael D.
Fisher, Karen

ORGANIZATION(S):

University of Washington, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Goals:

Community technology programs often lack the funding and technical assistance needed to sustain successful programs, let alone improve and expand their services. In the United States, these programs are training low-income youth, getting GEDs and technology skills to high school dropouts, upgrading the skills of recent immigrants, and working with a host of other residents to bring them towards success in the 21st century.

Internationally, community technology programs are often the first step in empowering individuals and communities to move toward fuller participation in a global society. However, ensuring digital inclusion requires effective policy and sufficient investment, and good research to determine impact and provide the necessary underpinnings for policy and investment decisions.

We are proposing a wildcard submission under the conference theme of "Community Technologies and Networking (CTN)," with participants from eight institutions heavily involved in CTN. Our wildcard session will comprise a 90 minute forum in which faculty and doctoral students—with varied academic and professional backgrounds—share their involvement, challenges and dreams for making a difference through CTN in different settings.

The anticipated outcome of this session will be a clear picture of the landscape for research programs moving forward, and a better understanding of the breadth and depth of work already being done or planned in this area across the iSchool community.

Building on the successful preconference community informatics workshop at the 2006 iSchool Conference at the University of Michigan, our wildcard session will be synergistic with the upcoming 2009 "Communities and Technology" Conference being held at Penn State. The wildcard session will provide an opportunity for the CTN community to meet and discuss multi-level issues with an aim of charting a unified path toward meaningful outcomes.

Format:

In preparation for the wildcard event, each participant will post the following information on our CTN WIKI on the iConference website (or another suitable location if this is not possible):

- 1) Affiliation, job title, project urls
- 2) A brief description of their involvement with community technology
- 3) Perceptions of the top 3 challenges in the area (e.g., funding for research, funding for practice, publishing/communication venues, policy conflicts, geographic, data synthesis, hardware/software, language, cultural...)
- 4) A brief description of their dream for the future

During the wildcard session itself, we will dedicate the first 30 minutes for each participant to provide a one minute overview of their primary research interests, in the context of the dreams and challenges previously submitted to the wiki (these will also be displayed in the room on flip charts for easy reference during the session). This will be followed by a 45 minute discussion period in which the participants and audiences will explore the interconnection of the challenges and dreams, identifying gaps or overlaps in the landscape.

The final 15 minutes of the session will be dedicated to a dot voting session, with each participant given 10 dots to be distributed across the challenges on the flip charts to end up with a consensus on areas that the research community might put the most effort into for the next few years. The summary results of this exercise will be posted to the session wiki after the conference, to encourage further discussion and evolution of ideas after the conference. We anticipate that this approach will provide all of us with some sense of where we might be able to work together to shape our future research, and a better perception of the field for everyone.

The following excerpts from participant input already received provide a good sense of the issues that will be explored in this session, all of which demonstrate the inter-disciplinary nature of this research area, and provide an excellent example of the richness that iSchools can bring to investigations of complex social problems.

- • Examine the context in which immigrants are using ICTs (at home, in community settings, etc.) and for what reasons, as well as the ways that ICTs contribute (or not) to participation and social inclusion.
- Gaining legitimacy for scholarship in this area
- Harnessing the interdisciplinary strengths of scholars and practitioners
- Developing theoretical frames that will continue to build scholarship
- Creating, within the iSchool movement, momentum for the application of the accumulating iSchool interdisciplinary knowledge to communities. Each school at present has developed parts of the ideal. I look forward to seeing well- developed community informatics programs in all schools, but we aren't there yet. This group can help move us toward that goal and build on the CI panels at the first and second iSchool conferences
- Fostering the sustainability of community networking projects
- Developing a clear research agenda (that emphasizes action or participatory research)
- Developing relevant coursework for the next generation of community network researchers
- Intellectual: Explore the utility of the concept of memory practices for online community research
- Blue sky: to secure funding for large scale research on community networks
- Explicitly thinking about how technologies can transform capacity building activities (that are non-technological) within communities
- Engaging communities to author, design, and develop infrastructures around these technologies
- Further understanding the conditions that lead to sustainable usages of technologies for capacity-building
- Further understanding economic principles that lead to community technology policy, so as to develop an agenda that can be transformative and aware of status quo agendas
- Lack of understanding of how digital inclusion policies can improve the business model for municipal wireless. Inclusion is frequently seen only as a cost.
- Funding. Community-based technology is essentially seen as "nice", but not vital to the functioning of society, to strategic research directions, etc.
- Defining the term "community" which has become quite problematic as it is increasingly applied to everything, and therefore increasingly means nothing.
- How recent (e.g., Web 2.0) technologies enable communities to form and how these new technologies encourage or inhibit behaviors and norms—especially ethical norms.

- The notion of communities of practice and how they affect the design and adoption of technical (especially information) systems.
- Going beyond the case studies that dominate this emerging field, either by strong theory building, larger datasets, or standardized methodology so our case studies can be compared more closely.

Participants:

As of the October 29, 2007 deadline, the following faculty and doctoral students have agreed to participate:

- Florida State University
- John Bertot (jcbertot@ci.fsu.edu)
- Charles McClure (cmclure@ci.fsu.edu)
 - Pennsylvania State University
- John Carroll (jcarroll@ist.psu.edu)
 - University of California at Los Angeles
- Ramesh Srinivasan (srinivasan@gseis.ucla.edu)
 - University of Illinois at Urbana-Champaign
- Caroline Haythornthwaite (haythorn@uiuc.edu)
- Kate Williams (katewill@uiuc.edu)
- Abdul Alkalimat (mcworter@uiuc.edu)
- Jon Gant (jongant@uiuc.edu)
 - University of Indiana
- Howard Rosenbaum (hrosenba@indiana.edu)
- Kathryn Clodfelter (kaclofde@indiana.edu)
 - University of Michigan
- Joan Durrance (durrance@umich.edu)
 - University of Toronto
- Nadia Caidi (nadia.caidi@utoronto.ca)
- Adam Fiser (adam.fiser@gmail.com)
- Danielle Allard (allard@fis.utoronto.ca)
- Diane Dechief (diane.dechief@utoronto.ca)
 - University of Washington
- Chris Coward (ccoward@u.washington.edu)
- Mike Crandall (mikecran@u.washington.edu)
- Karen Fisher (fisher@u.washington.edu)
- Karine Barzilai-Nahon (karineb@u.washington.edu)
- Bob Mason (rmmason@u.washington.edu)
- Chic Naumer (naumer@u.washington.edu)
- Carol Landry (cflandry@u.washington.edu)

We anticipate that a few more individuals may be added as the iSchool conference draws nearer.

WILDCARD SUBMISSION 14:
i-Schools and Undergraduate Education

AUTHOR(S):

Hansen, Derek L.

Bonzi, Susan M.

Kendall, Lori
McDonald, David W.

ORGANIZATION(S):

University of Maryland, United States of America

Syracuse University, United States of America

University of Illinois at Urbana-Champaign, United States of America

University of Washington, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Introduction

Undergraduate education is an important component of many i-Schools and of potential interest to many others. The nature of undergraduate education within the i-School movement will, no doubt, have a significant effect on our collective future. Undergraduate programs are often the most visible programs to other departments within the Universities in which we reside. They help shape first impressions of many current and prospective students. And, they force us to synthesize our various interests into a more basic core.

While all i-Schools have significant graduate programs, i-Schools vary widely in their stance toward undergraduate education. Some offer a full Bachelor Degree program, others offer a minor, and still others teach a single course at most. What's more, the undergraduate courses and curriculum vary widely across schools. This diversity is likely a natural outgrowth of historical processes unique to each University and is unlikely to disappear. Indeed, there is no one-size-fits-all approach to dealing with i-School undergraduate education.

However, despite our different needs, there are some fundamental issues that all i-School undergraduate programs must grapple with: the interdisciplinary nature of i-School education, the shared domain knowledge, and the need to explain our relatively new curriculum and programs to the wider University, students, and employers. This session is designed to start a conversation about these (and other) commonly shared issues and provide a platform for us to learn from one another's experience. It is most likely to benefit faculty members and administrators involved with undergraduate education in their schools. It is also intended to be useful for schools considering expanding their undergraduate program.

Goals

Specifically, the goals of the session are to:

- characterize the current state of undergraduate education in the i-Schools including how they fit within the wider Universities they are a part of;
- discuss common challenges and best practices for introducing and maintaining undergraduate programs in i-Schools, as well as the decision not to pursue an undergraduate program;
- characterize the core competencies that should be covered in i-School undergraduate programs; and
- help initiate relationships among individuals with an interest in undergraduate education.

Outcomes

If there is sufficient energy from participants, the session could lead to a brief report on undergraduate education within i-Schools. Depending on the nature of the report it could be shared via the i-School website or published in an appropriate forum such as the Chronicle for Higher Education or JALISE.

Format

The session will include an online component and a 1.5 hour session at the conference as described below. (Note: Ideally the session would last for 2 hours, but if that is not possible perhaps it could be held at a time slot where individuals who are still interested can meet together directly after the session for lunch or dinner).

Wiki Website

A publicly accessible wiki website will be created and used to compile information on current i-School undergraduate programs including links to their websites, names of current and past program chairs, and any documentation they are willing to share related to the programs. Those interested in attending the session will be encouraged to read through the materials before the session, in order to use our in-person time effectively. After the session, the wiki will be used to draft the report, which will summarize the discussion of the in-person session and summarize the documents collected on the wiki.

Conference Session

The conference session will approximately follow the schedule outlined below:

Introduction & Overview (20 minutes)

Participants will introduce themselves and their interest in undergraduate education. Derek Hansen (UMD) will give an overview of the state of undergraduate education within i-Schools based on the materials that were shared on the wiki website.

Prepared Presentations & Discussion (30 minutes)

A current or former undergraduate program chair of three different i-Schools will briefly discuss their program, focusing on its relationship with the wider University, its strategic position within the i-School itself, the core curriculum, and any major lessons learned. The three presenting i-Schools (UW, Syracuse, and ILL) were chosen because of their apparent success and their diversity of experience. Questions will be encouraged throughout the presentations.

Breakout Sessions (30 minutes)

At least two breakout sessions will convene after the prepared presentations. One group will focus on defining some of the core components of an i-School undergraduate curriculum. Another group will focus on the place of an undergraduate i-School in the broader University. Other topics may be proposed at the session itself. Each group will have a recorder who uploads a summary of their discussion to the wiki and a presenter who will briefly summarize their discussion to the larger group.

Wrap-Up (10 minutes)

We will summarize the discussion and agree on next steps, including the drafting of a report if there is sufficient interest.

Participants

The event will be organized by Derek L. Hansen, assistant professor of the University of Maryland. Like other i-Schools the University of Maryland is interested in expanding its undergraduate offerings and thus has an interest in this important topic.

The following participants have committed to share their school's experience:

- Susan M. Bonzi, Director of Syracuse's IST Undergraduate Program (1993-present) and Undergraduate Committee Chair (2002-present)
- Lori Kendall, UIUC Undergraduate Minor Committee Chair
- David W. McDonald, Former UW Undergraduate Informatics Program Chair (2004-2006)

WILDCARD SUBMISSION 15:

Can I-Schools Fulfill Their Mission Without Archival Studies?

AUTHOR(S):

Cox, Richard James

Larsen, Ronald

Cloonan, Michèle V.

Gilliland, Anne J.

ORGANIZATION(S):

University of Pittsburgh, United States of America

Simmons, United States of America

University of California, Los Angeles, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

If one listens in the library and archives communities to the whispers and rumors about the I-School movement, the sense one has about what is going on is that this is just another effort to move away from the traditional library school, merely another phase in the natural evolution of a portion of these schools into something far different than what the library science pioneers ever envisioned. Before the I-School movement was here, however, other aspects of these historic schools were evolving, such as can be seen in the history of graduate archival education having a place in the library and later library and information science programs. However, in this session we take a somewhat

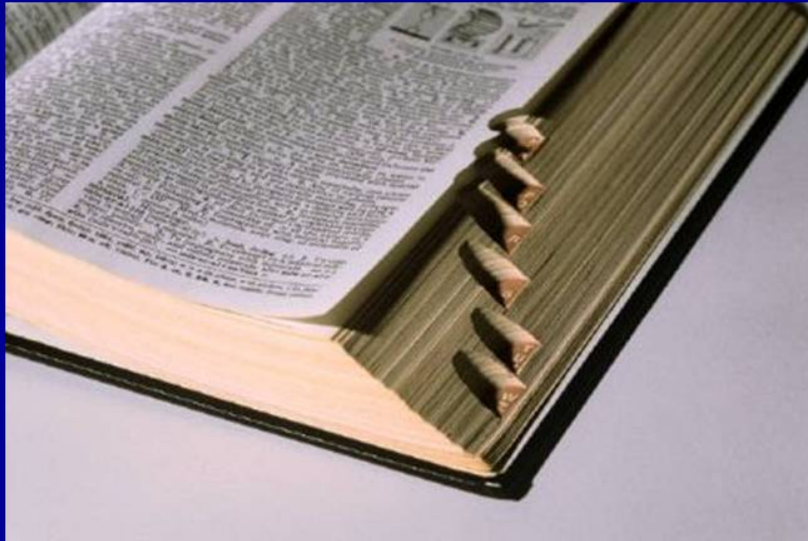
different tack, arguing that older areas of studies such as archival studies are essential to the future of I-Schools, particularly as these schools tackle the increasingly complex issues confronting a digital society. This session will consist of a preliminary paper co-written by leaders in archival studies and the concept of I-Schools (Richard J. Cox and Ron Larsen respectively) with two different commentators on the Cox-Larsen presentation who bring different perspectives to this topic. The paper will discuss the mission and history of I-Schools, the mission and history of archival studies, the basic elements and concepts of archival studies which are critical to I-Schools (curation and preservation of primary and secondary sources that are born-digital, the life cycle/continuum concept of records, the preservation imperative, cultural and humanistic perspectives, public and institutional memory, and the evolving notion of records in the digital era), and matters of the changing nature of personal and institutional archives which may be best fit into an I-School curriculum and mission (e.g., the implications of new portable digital technologies on issues such as the creation, maintenance, and use of records and information sources deemed to possess long-term archival value).

I-Schools & Archival Studies

Richard J. Cox & Ron Larsen

[Note: This is a working draft of a presentation for discussion; changes & revisions are expected before the conference]

Setting the Scene: Basic Definitions



Definitions: Archives

- Organizational, government, personal, or family records maintained because of continuing or enduring values
- Records preserved because of evidence, information, accountability, or corporate or public memory values
- Archives exist in every kind of organization -- in governments, corporations, libraries, universities, & cultural (museums, historical societies) organizations -- & are created & maintained by individuals & families

Questions: Who Are Archivists & What Do They Do?

- Are archivists part of the information professions? Or part of historical or cultural heritage fields?
- What is the nature of the knowledge domain of the archivist? How does it intersect with information sciences?
- How is the mission & work of the archivist evolving in light of digital recordkeeping & information systems?

Definitions: Preservation

- Maintain over time (usually considered as part of permanent value)
- Encompasses notions of conservation, restoration, and preservation management
- Aims for the maintenance of physical objects (artifacts) as well as for reformatting (such as digitization) for enhanced access & the long-term maintenance of digitally-born records & recordkeeping systems
- Preservation is generally seen to be the crux or end result of archival work (although archivists destroy more than they save)

Questions: Preservation in the Digital Era?

- How has the concept of preservation been challenged or transformed with the growing use of & dependence on digital systems?
- Are digital advocates still arguing that all information sources can be saved and effectively used?
- What is the ideal weighting between traditional and digital preservation in educating archivists (and preservation administrators)?

Definitions: Records

- Work of the archivist is centered about the identification, preservation, and providing access to “records” possessing archival value
- In North American practice, the notion of records was largely taken for granted, following general definitions created in government laws or best practices in corporate and other organizational settings
- Increasing use of information technology led to the need to revisit basic definitions and to re-engineer the uses of older archival sciences such as “diplomatics”

Questions: Is the Record Changing or Disappearing?

- Are new digital forms of records still functioning as transactions of business with the elements of warrant, structure, content, and context still relevant?
- Are researchers and others needing access to records still concerned about matters of authenticity and reliability as they once used to be?
- Are new means of providing access to more complex digital information sources trumping issues of definition and maintenance?
- Have the continuously emerging digital documentary forms eased the way for more postmodern notions of evidence and information?

Education & the Formation of Archival Knowledge



February 28, 2008

Cox-Larsen I-Conference Paper

9

Educating Archivists, 1930s-Present

- Apprenticeship, 1900-1960s
- Single Graduate Courses, 1930s-1970s, in history departments & library schools
- Three course sequence, 1970s- 1980s, mostly in LIS schools & first SAA guidelines & proliferation of institutes
- Emergence and decline of public history programs, 1970s-early 1990s
- Employment of Regular Faculty, 1980s to present, & new SAA guidelines in 1988 & 1994
- Existence of Multiple Faculty programs, mostly in LIS schools, 1990s to present
- Beginning of doctoral programs, 1990s to present

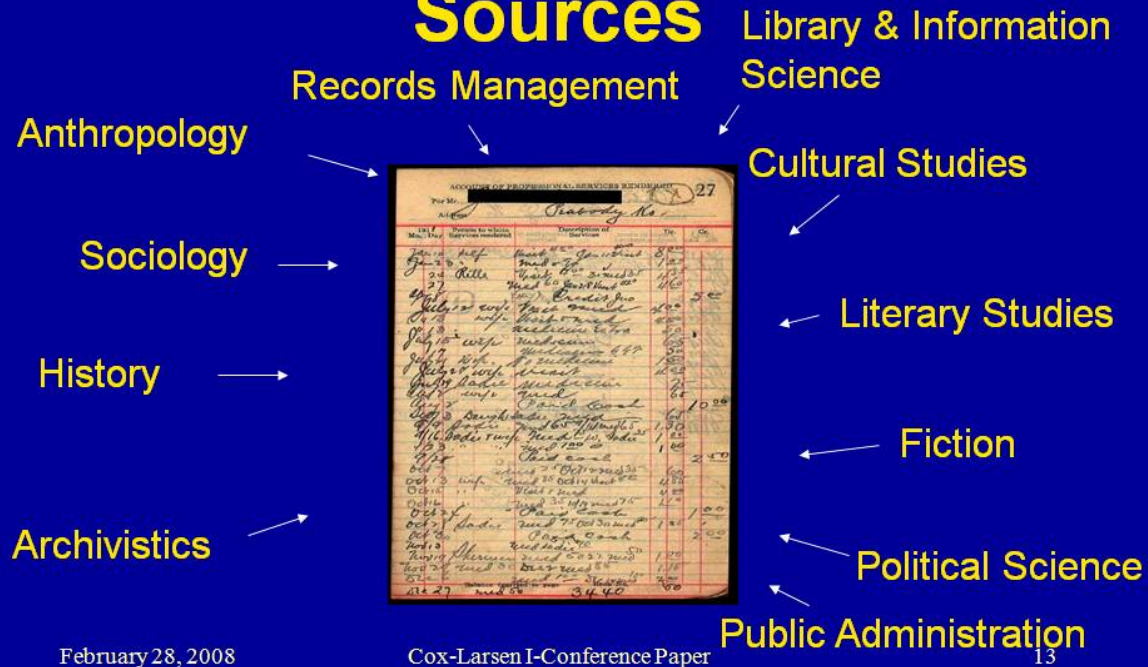
Key Features of Graduate Archival Education

- Generally focused on traditional records systems & archival principles built on such systems
- Some stress on electronic records management issues, usually presented either in dedicated course or integrated throughout curriculum
- Limitations posed by small faculties, adjunct reliance, nature of area archives, & other similar factors
- Many students attracted by aim of working in cultural organizations with older, traditional forms of documents

Expanding Notion of Archival Knowledge & Scholarship

- Scholars -- from literary & cultural studies, anthropology, history, sociology, political science, & other fields -- are studying archives or the "archive" & adding new understanding to archival studies
- New forms of scholarship -- embracing digital means of collaboration & access -- are suggesting new uses of archives (both digitally born & digitized)
- Recordkeeping & the scholarship on it represents a "relatively new field of study. The boundaries of the field are poorly defined and porous. This is characteristic of emerging disciplines and need not be a cause of professional insecurity."

Archival Knowledge Sources



February 28, 2008

Cox-Larsen I-Conference Paper

Public Administration

13

Transitioning to a Digital Era

- Mark Herring: “if we define *knowledge* as any bit of datum, right or wrong, factual or not, fraudulent or accurate,” then the digital world is fine, but “if this is the definition of information that we want, then, yes, the Web should replace all libraries. On the other hand, if knowledge includes something about accuracy, appropriateness, balance and value then the Web cannot arrogate to itself a place of preeminence to knowledge-seekers.”
- Jeff Gomez: “And so to expect future generations to be satisfied with printed books is like expecting the Blackberry users of today to start communicating by writing letters, stuffing envelopes and licking stamps.”

Challenge of ERM & Digital Archives

- “Most organizations have serious operational shortfalls regarding the processes by which they manage electronic records, one of their most important assets” Robert F. Williams and Lori J. Ashley
- “As we face the challenges of electronic records, we must also face our need for new knowledge. We need new tools for new materials. Where to begin?” Richard Pearce Moses
- “While we are still at the dawn of the digital era, before too many cultural assets are lost, and before the technology has raced utterly beyond our ability to catch up, we need to construct concepts, methods and operational systems that can preserve and provide access to digital information.” Ken Thibodeau

What Could Be Gained by Strengthening Archival Studies in I-Schools



Defining I-Schools

- Focus on relationship between information, technology, and people.
- Commitment to learning and understanding the role of information in human endeavors.
- Conviction that expertise in all forms of information is required for progress in science, business, education, and culture.
 - Uses and users of information
 - Information technologies and their applications.

Regaining the Ancient Model

- In ancient world, the “scribes were not merely penman and copyists but intellectuals,” but the “academics of their time” In Israel, scribes were part of an exclusive group: “The skills of the scribes – of reading, understanding, and interpreting – commanded general respect. The scribes held the key to the symbolic capital of the nation” Karel Van Der Toorn
- In modern world, “A competent archivist is to be looked upon as a scholarly colleague of the researcher, far more than solely a preserver and a caretaker. His knowledge of the sources can contribute materially to the user’s evaluation and understanding of them.” Philip Brooks

Archives & the Liberation of Scholarship

- “Digital archives combined with new technologies will liberalize scholarship. They will enable simultaneous access to a range of sources (both local and distant) and facilitate the use of research methods not possible with conventionally printed or hand written records.” Seamus Ross
- And this can only occur in a new collaborative environment: “No one can work in isolation on digital preservation and access issues because the needs and requirements are too great. We all benefit from (and generate) economies of scale, pooled expertise, larger funding, and more robust infrastructure when we collaborate. And collaboration means not just crossing over our museum/library/archives divisions, but entering whole new communities such as science, engineering, and the commercial sector.” Diane Zorich

What I- Schools Offer to Archives That LIS Schools Do Not

- **Holistic perspective**
 - Principles, traditions, values
 - Technologies, standards, policies
 - Society, community, individual
- **Interdisciplinary linkage**
 - Explicit collaborative relationships
 - Joint degrees
 - Minors
- **New image**
 - Defining the future
 - Transitioning the cultural and societal values of the past

Emerging challenges

- **Geometric growth of information**
 - 161 x 10⁹ GB in 2006
 - 40% / year growth rate
 - 1 online + 3 tape backups = \$7T / yr (~ national debt)
 - “Humans can read ~2GB in a lifetime” Rhys Francis, CSIRO, AU
- **Slow uptake of institutional repositories**
- **Support for data-driven science**
 - Computer analysis of vast stores of information
 - “When collections get large, only the computer reads every word.” Greg Crane, Tufts
- **Impact of mass digitization projects**
 - Google Books
 - Million Books Project

Emerging challenges, cont'd

- Capturing content at its source
 - Original data
 - Metadata at point of creation
 - Usage profiles & annotations of users
- Organizing vast content for access and use
 - By humans
 - By computers
 - Supporting the evolution of relationships among users, data, information repositories, and communities
- Moving beyond recognizing the inevitable to dealing with it
 - Pervasive digitization
 - Ubiquitous access
 - At-risk content

Unfolding Opportunities

- Cyberinfrastructure initiatives
 - Recognizing “content” as infrastructure
- Definition of value-added services
 - Cataloging discrete objects within collections
 - Named entity services
 - Customization and personalization
 - Structured user contributions
- Curation beyond the artifact... data journals
 - Peer-review systems built around data sources
 - Coherence, design, consistency, reliability of access, ...
- Leveraging the lessons of innovative projects
 - Human Genome Project
 - National Virtual Observatory
 - Physics and Chemistry

Expanding Domains of Concern

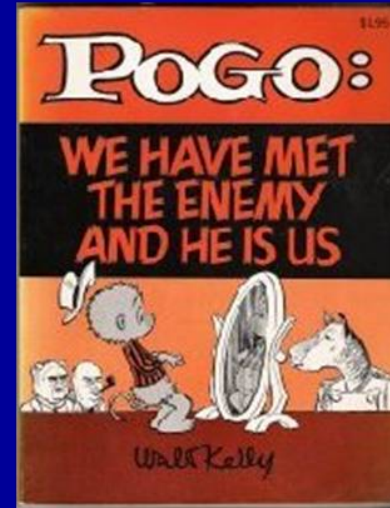
- Variety
 - Differences in scale, structure, complexity, ...
 - Differences in privacy, copyright, standardization, commercial value, ...
 - Among traditions of scholarly communities
- Resistance to change
 - Adaptation of organizations, traditions, incentives, policies, economics, law, ...
- Managing scale and complexity
 - Locus of responsibility: the academy vs. industry
 - The role of supercomputing centers (superdata centers?)
- Access
 - Publications, models, simulations, software, data, ...
 - Image, film, sound, statistical data, ...
 - Restrictive IPR and business practices
- Stability
 - Infrastructure, including organizations, technology, and data
 - Federal aversion to long-term commitments

What Archival Studies Offers to I-Schools

- Vision
 - We've been this way before
- Values
 - It's the evidence of our very existence
- Models
 - Ways to think more broadly about our discipline
- Case studies
 - Horror stories vs. Hero stories
- Perspective
 - Sometimes it's ok (even necessary) to throw it away
- Grounding
 - Digital or analog, it's all information

Concluding Thoughts

- Are we part of the solution?
- ... or part of the problem?



February 28, 2008

Cox-Larsen I-Conference Paper

26

WILDCARD SUBMISSION 16:

How public is the Internet? A conversation on the nature of human interactions on-line and the implications for research methods.

AUTHOR(S):

DeLap, Alpha Selene

Edwards, Phillip M.

Organization(s):

ORGANIZATION(S):

University of Washington, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

In the past decade, “virtual” research—empirical investigations conducted via the Internet—has increased dramatically across a variety of disparate disciplines. Areas such as cybersecurity and encryption, digital government and citizenship, consumer health informatics, and user behavior in online spaces have emerged to become signature iSchool research areas, often shared with particular disciplinary heritages (e.g., computer science, political science and communication, public health, and sociology—respectively, but not exclusively). In addition, the field of Information Science is dominated by research developing or using emerging technologies. These new technologies often occupy a gray area in which ethical issues either have not been sufficiently well-defined or push against existing definitions. Questions surrounding the “public” nature of the internet and Web 2.0-era information technologies have also emerged and have become increasingly urgent given the tightening of federal, state and University regulations as they relate to the protection of human subjects.

At the convergence of multiple disciplinary and methodological perspectives, Information Science researchers are well-positioned to become more active participants in both scholarly and institutional conversations regarding the

appropriate risks and benefits that participants in online research studies might be subject to. Critiques of IRB inconsistencies exist, what we need is a thoughtful and thorough community response to the innately complex nature of virtual research and a map which can guide us towards the future and the study of twenty-first century systems, selves, and societies.

Our goal for this wildcard event, is to generate a lively and rigorous debate which accomplishes the following three goals: 1) extends the dialogue within the Information Science field concerning the beneficence and respect for participants in online research; 2) enumerates a set of best practices for iSchool researchers in relation to conducting approved research on-line and; 3) moves us towards the process of drafting an iSchool set of ethical guidelines related to virtual research.

WILDCARD SUBMISSION 17:

What to do with Geeks and Nerds? A Collaborative Website Analysis Workshop

AUTHOR(S):

Stam, Kathryn
Guzman, Indira Rita
Scialdone, Michael

ORGANIZATION(S):

SUNY Institute of Technology, United States of America
TUI University, United States of America;
Syracuse University, United States of America

ABSTRACT: (SEE THE FOLLOWING PAGE IF ABSTRACT DOES NOT APPEAR BELOW)

Wildcard proposal: What to do with Geeks and Nerds? A Collaborative Website Analysis Workshop

Social scientists no longer have to travel far and wide for data on social life, as it is now available at our fingertips on the web. However, many of the techniques normally used in traditional research such as participant-observation and the writing of field notes, need to be adapted quite a bit for use in this new media. Rather than data being expensive and precious, the dilemma becomes how to choose from so much information and how to store and analyze it in meaningful ways that remain acceptable within more traditional fields of study.

In this interactive workshop, we will start by introducing our previous work on IT occupational culture, the stereotypes of geeks and nerds, and the potential implications of these stereotypes in the recruiting of people to information technology careers. Taking the topic of geeks and nerds in the IT culture as the topic for our web research, we will describe our experiences on conducting web research by archiving websites using the Mozilla Firefox scrapbook Add-on tool. We will talk about the tool and its features and briefly provide examples of previous research explorations. Then, we will present examples of our research journey on our topic using the scrapbook tool collecting and analyzing data from several websites (commercial sites, blogs, videos) as an experiment to see what we can learn about the culture of geeks and nerds by viewing different types of websites. Finally, as the heart of the workshop, we will have a collective activity of brainstorming, applying, and comparing techniques for webpage analysis, using the topic of geeks and nerds in the IT culture as subject material for discussion. We plan to form multidisciplinary groups from the audience and facilitate a half-hour break-out session to allow the smaller groups to share ideas and summarize their initial findings. The groups will then reunite for a large-group exchange about archiving and analysis of websites about a certain theme.

The organizers of this workshop embody a multi-disciplinary approach in all of our work and look forward to the opportunity to forge diverse groups to help analyze the websites from different perspectives and inspire new thinking about possible methods for website analysis.

Indira Guzman is an Assistant Professor in the College of Information Systems and the College of Business Administration at TUI University. With a background as a chief and administrator of information systems, she has conducted research in human resource in IT, career orientations, IT occupational culture, gender and ethnic diversity, recruitment, and retention. She obtained her PhD from the School of Information Studies at Syracuse University. Kathryn Stam is an Assistant Professor of Anthropology at the SUNY Institute of Technology, and has

brought her ethnographic experience to her study of the effects of technology on social relationships in the workplace and within the family. She and SU's Jeff Stanton recently published a book called *The Visible Employee: using workplace monitoring and surveillance to protect information assets- without compromising employee privacy or trust*. Michael Scialdone is a doctoral student from the School of Information Studies at Syracuse University with experience, conference presentations and publications related to ethnography of massive multi-player online role-playing games (MMORPG's) including Furcadia and Runescape.

Ten Tips for Improved Presentations

Ron Dietel

UCLA Graduate School of Education & Information Studies

Recommendations presented at the 2008 iConference: *iFutures: Systems, Selves, Society*

- 1) The **best way** to improve your presentation is to **practice it at least 3-4 times** before you give it. Repeated practicing will reduce jitters, improve timing, help you make effective transitions, and increase your confidence. Practice in front of a mirror or better, videotape yourself.
- 2) **Less is more.** When we try to cover too much content or use too many visuals, we speed up our pace to an uncomfortable level. Reduce both content and visuals so that you have time to emphasize the most important points, leaving the details to a paper, article, or to the Q&A part of the session.
- 3) **Own the room.** A presentation is your opportunity to shine. Don't refer to other presenters on your session unless connecting a point between that presentation and yours. Words like "that's a tough act to follow" or "how much time do I have left" tell the audience that you either lack confidence or have not practiced.
- 4) **Study** effective speakers and **steal** every good thing they do. Good speakers for example, pause for effectiveness, vary their voice levels to emphasize key points and keep attention, and often integrate small bits of humor into their speech. Most importantly, they have practiced their presentation many times.
- 5) **Smile and make eye contact.** Nothing warms up an audience like a big smile at the beginning of your presentation. Work your eyes around the room, with a goal of making direct contact with each person in the audience at least once.
- 6) Arrive at your presentation site a **minimum of one hour** before your session begins. Close the blinds, adjust the heat, check the microphone and make sure that logistic issues don't detract from your presentation. If the room is vacant before your presentation, practice your presentation to an empty audience. You will feel more comfortable and confident.
- 7) **Avoid going last** in a presentation, especially if there are four presenters or three presenters and a discussant. Far too often, the last person's time gets cut short because of poor time-keeping, discourteous presenters who speak over his or her time limit, or audiovisual problems.
- 8) Use the **666 rule** for PowerPoints or slides: no more than 6 bullets on a slide, no more than six words per bullet, and no more than 6 information points on a graphic. Putting too much text or information on a visual, often results in reading the presentation from the slide vs. making an effective speech to your audience.
- 9) **Hold the handouts** until the end of the session. If you give out handouts before the presentation begins, people will know what you are going to say before you say it and pay more attention to the handouts than to you.

- 10) If you can only manage one of the above, make sure it is number one. How do you get to Carnegie Hall? **Practice, practice, practice.**

Top Ten Tips for Media Relations

Marlo Welshons, Assistant Dean for Publications & Communications
University of Illinois Graduate School of Library and Information Science
Recommendations presented at the 2008 iConference at UCLA

- 1) **Identify the communications staff person at your department or college.** If your unit doesn't have one, get to know the university communications and/or news bureau staff. Establishing a good relationship with those staff members will be key to their ability to understand your work and therefore write about and promote it.
- 2) **Identify your specific areas of expertise,** and make sure the communications staff member knows what these are. Expertise can be different than your current research interests: areas of expertise are those on which you could speak with authority to a reporter.
- 3) Remember that **it is never too early to give communications staff a heads-up** about an upcoming event or news story. It can, however, be too late.
- 4) **Make a list of niche publications** for distribution of press releases or placement of stories — communications staffers are not always aware of all relevant journals or magazines in a particular field (especially if the staff member works for the campus news bureau, not your specific academic unit).
- 5) Remember that when determining whether or not to write a story, communications staff and reporters will need to **assess the newsworthiness of the story.** In other words, they will ask, "so what?" Be prepared to explain the impact of your research — in terms that non-specialists will understand — by writing concise answers to the following questions: Why is your work important? What makes your contribution unusual? Who will benefit and how? What is the most important point to make?
- 6) **Respond promptly to media requests,** even if only to decline an interview. Reporters usually are on a deadline; a delay of a day, or even an hour, can mean a lost opportunity, and perhaps a reporter disinclined to call you or your university again.
- 7) If a reporter contacts you unexpectedly, **ask questions** about the type of story the reporter is pursuing, the context in which you might be quoted, and the background of the reporter and the publication or broadcast outlet. Feel free to ask for clarification on any of their questions, and don't limit yourself to responding to those questions; emphasize points you feel are important.
- 8) **Remember the audience:** Keep statements clear and concise. Provide plain-language interpretations and metaphors. If you don't, the reporter may choose language that changes the meaning of the story.
- 9) Assume that **everything you say will be quoted.** If you feel that making a comment would be inappropriate or outside your expertise, politely decline. Avoid saying, "no comment," which usually sounds defensive; instead, offer to check on the requested information and get back to the reporter before his or her deadline or offer the names of others who may be in a better position to answer their question.
- 10) **Don't ask to review or approve the article.** Do ask when the article will appear and immediately inform your unit's communications staff about any contact with the media — they may be able to follow up with the reporter to provide a press kit or additional information.

Thanks to the University of Illinois Office of Public Affairs and Ron Dietel for their help in producing this handout.



Information. The Power to Transform Our World.

Writing Tips for Academics

Kelly Shaffer
Director of External Relations
University of Pittsburgh
School of Information Sciences

Thanks to Ron Dietel, Marlo Welshons, and Charles DuBois for their help in producing this presentation.



General Guidelines

- Carefully define your question/argument first
- Consider your audience – journal, grant, conference
- Assess the knowledge level of your audience
- Draft an outline -- and write to it!
- Proofread your work 24 hours later!
- Use a professional – if there's time.



Information. The Power to Transform Our World.

What is the question or problem?

- Identify – what is it?
- Justify – why is it important?
- Develop a thesis statement that outlines your approach



Who is the audience?

What is their knowledge level?

- Journal – very knowledgeable about field
- Conference – less familiar with your field
- Grant – initial reviewer might have passing familiarity with field
- Media – assume no knowledge of the field

Your audience determines your tone and your language.



Draft an Outline

- Start with thesis statement or statement of problem
- Carefully note critical points to make
- Write your conclusion statement
- Track your references as you write
(you won't be able to find them later)



Proofread Your Work

- Did you....
 - Clearly define a problem or statement?
 - Follow your outline?
 - Maintain a style in terms of verb tense and language usage?
 - Cite appropriately?



Information. The Power to Transform Our World.

Use a professional

- If you need it, hire a proof reader or editor, to review your work
- Often, your university or iSchool has such a resource for you!



Tips from an academic editor

- Don't cut and paste without careful review.
- Don't wait to start.
- Save copies of reference materials.
- Talk to your iSchool communications liaison to see what resources are available.
- Let your audience dictate your style.

UCLA
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Presentation Tips for All of Us

Ronald Dietel

UCLA Graduate School of Education & Information Studies

February 29, 2008

Overview

- Design
- Delivery
- Environment

Design - Content

- Research Questions
 - *Are the district's formative assessments aligned to state and school district content standards?*
 - *Do the district formative assessments improve learning in mathematics?*
 - *What is the correlation between performance on the district assessments and performance on the state test?*
- Formative assessments that are not aligned to standards are not useful in guiding and informing instruction, Harker and James, 1998
- Analysis of mathematics content standards that had been used to design third-grade assessments, was then used to design instruction—with learning experts
- Random selection of 400 students who received different types of math instruction (ten lessons each) before taking the district mathematics assessment:
 - *Data on mathematics alignment*
 - *Evidence collected in mathematics formative assessment using different mathematics curricula and instruction*
 - *Learning experts analysis of data*
 - *Triangulation of data*
 - *Study Limitations*

- ## Focused Content



Design - Key Points

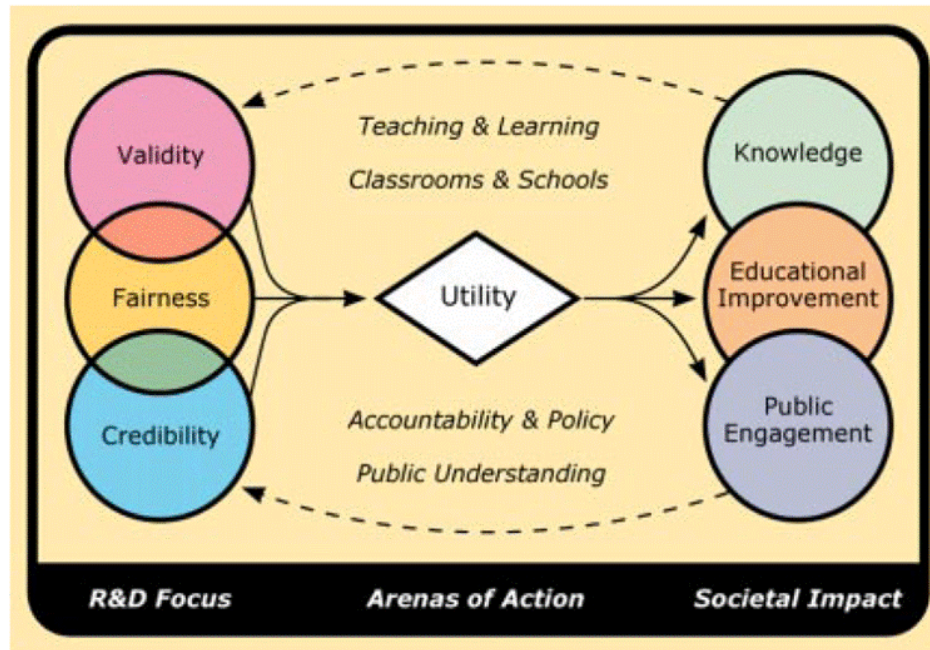
- **Formative assessments that are not aligned to standards are not useful in guiding and informing instruction, Harker and James, 1998**
- **Analysis of mathematics content standards that had been used to design third-grade assessments, was then used to design instruction—with learning experts**
- **Random selection of 400 students who received different types of math instruction (ten lessons each) before taking the district mathematics assessment**
 - *Data on mathematics alignment*
 - *Evidence collected in mathematics formative assessment using different mathematics curricula and instruction*
 - *Learning experts analysis of data*
 - *Triangulation of data*
 - *Study Limitations*

Focused Key Point

Our study found...

- District formative assessments had little influence on learning

Validity, Fairness, Credibility, Utility Are Underlying Principles



Design - Weak Visuals

- Cover too much content
- Too much content on one slide
- Difficult to read

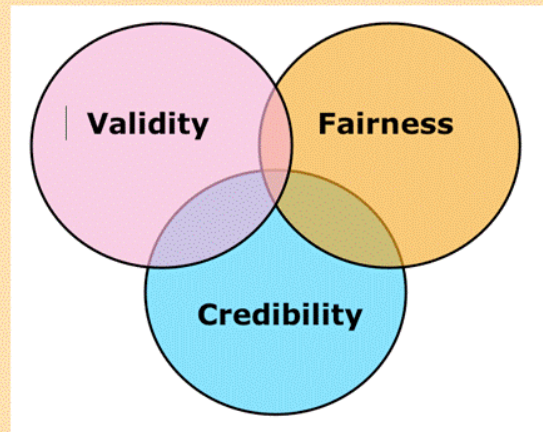
UCLA

Better Visuals

Follow the 666 rule...

- 6 or less bullets per slide
- 6 or less words per bullet
- 6 or < information points on graphics

Fewer Information Points



UCLA

Common Delivery Problems

- Reading paper or slides
- Nervousness
- Poor eye contact
- Difficult to understand
- Lack of self-confidence

Improving Delivery

- Learn from others
- No excuses
- Eye contact; gestures
- Time-keeping
- Hold the handouts
- Carnegie Hall

UCLA

Improving Environment

- Plan in advance
- Arrive early
 - *Visuals*
 - *Sound*
 - *Lighting*
 - *Temperature*

10 Tips for Great Presentations

1. Practice
2. Reduce content
3. Own the room
4. Learn from others
5. Eye contact

UCLA

10 Tips for Great Presentations

6. Arrive early
7. Avoid being last
8. 666 rule
9. Hold handouts
10. Practice some more



Media Relations for Academics

Marlo Welshons
Assistant Dean for Publications & Communications
University of Illinois
Graduate School of Library and Information Science

Thanks to the University of Illinois Office of Public Affairs and Ron Dietel
for their help in producing this presentation.



Information. The Power to Transform Our World.

General Guidelines

- Identify your communications liaison.
- Identify your areas of expertise.
- Keep your liaison informed—as early as possible.
- Identify specialized publications in your field.
- Prepare to justify newsworthiness of your story.



Preparing for an Interview

- Respond promptly to media requests
- Ask about:
 - the type of story the reporter is pursuing
 - the context in which you might be quoted
 - the reporter's and publication's background



Information. The Power to Transform Our World.

Preparing for an Interview

- Consider these questions:

- Why is your work important?
- What makes your contribution unusual?
- Who will benefit and how?
- What is the most important point to make?

- Write out answers in advance.



During the Interview

- **Assume everything you say will be quoted.**
- **Avoid “no comment.”**
- **Don’t speak beyond your expertise.**
- **Provide support where possible; offer to promptly send background material.**



Information. The Power to Transform Our World.

During the Interview

- Listen carefully to questions, ask for clarification.
- Don't just respond to questions; emphasize important points.
- Remember audience: use plain-language interpretations and metaphors.



After the Interview

- **Don't ask to review or approve the article.**
- **Do ask when the article will appear.**
- **Follow up with additional information/clarification.**
- **Notify your communications liaison.**
- **Know that not every interview will result in a quote.**

Top Ten Tips for Improving Your Writing

By Kelly Shaffer, the iSchool at the University of Pittsburgh

With contributions from Ron Dietel, CRESST at UCLA

1. State your question, problem or thesis clearly and concisely in the beginning.

This will grab the attention of your audience and let them know what you are trying to explain quickly and succinctly. In addition, it will encourage them to read on if they have an interest in the subject. This introduction is your chance to “sell” the reader.

2. Start with an outline.

If you craft an outline, you will see how to logically present your argument. You will also be aware of the critical points that must be made, and in the proper order.

3. Use the active voice, rather than the passive.

This will result in a livelier article that is more interesting to read. It also results in stronger, clearer sentences.

4. Avoid jargon.

This is particularly important if you are writing for an audience which is unfamiliar with your topic. If you must use jargon, explain it clearly the first time the word or phrase occurs in your piece.

5. Proofread carefully.

Since the advent of “cut-and-paste,” we often reuse material that might have errors or inappropriate verb tenses.

6. Review pieces by multiple authors carefully.

When several people contribute to the same article, check the finished work carefully for stylistic differences, repetitive phrases or ideas. Pay careful attention to ensure similar sentence structures and verb tenses.

7. Have someone else proofread your work.

Your eye will skim over words that you have written. You will miss details, or become so caught up in the details that you miss the point where you deviate from your structured outline.

8. Double-check your references.

Make sure that each citation is in the proper format. Check to make sure that each citation has an entry in the references section; conversely, check to make sure that each listing in the references section matches a citation in the work.

9. Know your audience.

Writing for students is different than writing for professionals in your field. Writing for the media or non-domain readers is another matter completely. You must tailor your writing to the knowledge level of your audience. In addition, consider how the reader will view your work – on a computer screen, a projection screen, or an actual piece of paper. The resolution of graphics needs to be adjusted to meet the viewing format.

10. Volunteer as a reviewer.

If possible, volunteer to serve as a reviewer for a conference or journal. Exposing yourself to writing by others will help you to refine your own style.