Cultural Heritage Information Dashboards

Richard J. Urban
Graduate School of Library and
Information Science
University of Illinois
Champaign, Il

rjurban@illinois.edu

Piotr Adamczyk
Metropolitan Museum of Art
1000 Fifth Ave.
New York, New York
piotr.adamczyk@metmuseum.org

Michael B. Twidale
Graduate School of Library &
Information Science
University of Illinois
Champaign, Il

twidale@illinois.edu

ABSTRACT

Large-scale aggregations of digital collections from libraries, archives and museums offer users unprecedented access to cultural heritage materials. But they also have failed to incorporate important contextual information that allows users to develop an understanding of the significant features of purpose-built collections. This paper explores the development of information dashboard prototypes that provide users a high-level overview of cultural heritage collections. Two case studies using rapid-prototyping methodologies are presented.

Categories and Subject Descriptors

D 2.1 Requirements/Specifications [Elicitation methods]

General Terms

Design, Human Factors

Keywords

Rapid prototyping, information visualization, dashboard, cultural heritage

1. Introduction

Currently available digital collections from libraries, archives and museums represent just the tip of the iceberg – millions more analog resources could potentially be added in the next decade. As these resources move from physical repositories into networked environments they run the risk of losing connections to institutional and curatorial contexts that make "collections" significant as curated, identifiable wholes. Traditional approaches to description have relied on textual metadata augmented by the tacit knowledge of the professionals who care for collections. In order to gain access to physical collections it was necessary to work through local institutional infrastructures [[1], [2]]. Although institutional websites may recapture some of the essence of this infrastructure, many of the subtle affordances of physical modes of access have not carried forward into the digital realm.

While traditional modes of collection access may have been sufficient for distributed institution based materials, they also suffered from the known problem of "collections understanding" among both professionals and users. Lee discovered that the meaning of "collection" is difficult to pin down even among librarians, archivists and museum professionals who work with them every day. Professional perspectives of how collections are managed diverge from the user's perspective of how collections are used, but it is the professional perspective that has largely shaped collection-level descriptive practices – often without a clear grasp of the role that collections play in information seeking behaviors [3]. Professional practices for collection-level and item-level metadata have also proceeded independently of each other without consideration of how thy could be mutually supportive. The problems of "collection understanding" are further exacerbated by the rapid changes that online access to cultural heritage materials brings. Large-scale aggregations containing collections from multiple institutions add to the problem and may make individual items "informationally small" due to the lack of collection-level metadata [4]. Many current descriptive practices are based on an assumption that the institution is the gateway. However usage logs show that users increasingly arrive at cultural heritage materials not through the front door, but through broader information seeking behaviors on the open web that drop them into the middle of a collection or aggregation. In such cases, users may not be able to discern who owns a collection, how big it is, what it contains, its relative strengths compared to other collections and the significance of its constituent items.

Fortunately, lower barriers to information visualization are creating opportunities to re-imagine how we represent the contours of collections. Our research seeks to restore the contextual information that digital collections have lost through the use of "collection dashboards." Using a user-centered rapid-prototyping method known as "patchwork prototyping" we elicited initial concepts from humanities scholars and library, archive and museum professionals in conference settings. These concepts served as the basis for two prototypes - one which uses OAI-PMH Dublin Core metadata aggregated by the IMLS Digital Collections and Content (IMLS DCC) Project and the other using publicly available data about the Metropolitan Museum of Art's collections. In both cases, the quality of metadata syntax and semantics created significant barriers to successfully automating the generation of dashboards. We discuss a number of issues raised by the initial prototypes.

2. Background
2.1 Visualizing Cultural Heritage Collections

As the power of graphics processing has increased, the ability to visualize larger and more complex data sets increasingly has come within the reach of the cultural heritage sector. This lower threshold, combined with increasing amounts of metadata about cultural heritage materials, creates an opportunity to use exploratory information visualizations to support information seeking activities. The research literature on information visualization is both broad and deep, but has largely been confined to the lab, beyond the reach of many libraries, archives and museums. The work of Martin Wattenberg, Fernanda Viégas

and others interested in "democratizing" information visualization has made these techniques publicly available through web services such as Many Eyes™ and the Google Visualization API [5]. Visualization tools have left the lab and entered the mainstream through interactive sites such as the New York Times Visualization Lab, representations of personal social networks or previously incomprehensible government data. [6], [7], [8].

Like many of these examples, libraries, archives and museums posses a treasure trove of complex data describing their collections. Traditionally this information is put in the service of text-based search and retrieval systems, but these new and more broadly available information visualization techniques offer the opportunity to re-imagine how this information could be used or improved to facilitate such visualizations. New publicly available visualization approaches can benefit from the foundational research in visualizing cultural heritage collections. Green, et al. developed a novel browser for Library of Congress American Memory collections that included timelines, topical browsers and geospatial views of collections [9]. Also using the American Memory collections, Derthick's *Bungee View* presents a faceted browsing interface that also highlights the relative size of facets in a collection [10]. Using archival descriptions, the ArchivesZ project and the *Visible Archives* attempt to provide users a sense of the relative size and scale of archival series [11], [12].

A related area is the development of "information dashboards" that may include multiple information visualizations. The dashboard metaphor has its roots in the business sector where they are used to track large, complex and changing data from financial markets, industrial processes or inventory flow [13]. A few libraries have adopted the dashboard metaphor to make circulation and reference statistics available to managers – or the public, such as George Legrady's *Invisible Made Visible* installation at the Seattle Public Library [14]. Dashboards have also made inroads to the cultural heritage sector through financial products that allow non-profits to view endowments and investments. The Indianapolis Museum of Art (IMA) has adopted the dashboard metaphor to provide visitors with information about attendance statistics, power consumption and what is blooming in the garden this week [16]. Part of the aim of the IMA's dashboard is to make this information more public – a kind of "radical transparency" whereby an organization chooses to share data as a means to include a wider constituency in its core processes [17]. For a museum such as the IMA, this can be a kind of public outreach or recruitment. The data to be shared may not on its own be very exciting, but when combined and visualized it can tell a compelling story about the contributions of an institution to a community. This exploration of collections dashboards seeks to extend this kind of transparency to the collections themselves.

2.2 Patchwork Prototyping
In order to explore the problem space of collections dashboards, this research adapted the patchwork prototyping approach [17]. This rapid-prototyping method combines user-centered design with high-fidelity prototypes constructed using open-source software and freely available web services. While this "mashup" approach has been used to generate interesting web services (such as a Google Map of Craigslist apartment listings), patchwork prototyping explicitly ties them to prototyping tasks. Because it relies on "off-the-shelf" components loosely stitched together, patchwork prototyping can fills the gap between low and high fidelity prototypes.

By itself, patchwork prototyping does not address the common problem of access to intended users of novel products. The number of people willing and able to help may be limited, especially if a certain level of skill is required, and the time they can devote to helping is also limited. In the case of this project, our target users are professional historians, researchers and skilled amateurs. Additionally librarians, archivists and museum professionals because of their intermediary experiences can also inform our design. Along with this informant constraint common to many projects, we wanted to try to generate and exploit new opportunities for informing design. How might target groups contribute in very fast, lightweight, low commitment ways, other than more conventional participatory design meetings that typically take at least an hour? One idea was to do some of our design work at conferences. Many conferences have a demo session where a working application or prototype is shown off. Attendees typically offer informal comments and although these can and do inform the designers, this does not seem to be considered as a design process. What if instead of doing a conventional demo of a finished product, the demonstration session was used as a platform for participatory design?

When demonstrations are used as a user-centered-design (UCD) process rather than a conventional demo, it can have very desirable properties. Demonstrations clearly simplify recruitment. The process of participating / volunteering is less effortful, one might even say less aggressive than asking a person to help in a conventional UCD session. Interactions are understood to be typically short, typically 5-10 minutes, lowering the perceived time commitment. Participation can be incremental, starting from a peripheral observation and calling out questions or comments, gradually moving up to hands-on interaction. It is contextually acceptable to participate for a short while and then move on to another demo. A demo session is noisy and messy, with considerable movement affording multiple conversations and many kinds of interactions. Finally it is very easy for a participant to politely disengage and go to another booth whenever they want to. Ease of disengagement helps in encouraging initial engagement because the level of implied commitment is less.

3. Case Studies
3.1 Opening History
The Institute for Museum and Library Services Digital Collections and Content (IMLS-DCC) Project began in 2002 in order to aggregate collection-level and item-level descriptions from digitization projects funded by the Institute of Museum and Library Services (IMLS). Beginning in 2007, the IMLS DCC project expanded the scope of its aggregation activates to digital collections related to American history. Known as Opening History this aggregation currently contains more than 770 collection-level descriptions and over 950,000 item-level descriptions from more than 300 different libraries, archives and museums (LAMs) [18]. Because of the diversity of materials aggregated, we have had a keen interest in the role that collections play in users' ability to identify useful resources [[19], [20]. An important part of our current workplan is exploring how to better represent the content and contexts of Opening History for scholarly use and users.

Previous development work for IMLS DCC/Opening History has followed a traditional approach of using low-fidelity paper prototyping to generate design ideas along with high-fidelity working prototypes that were used for usability testing. While the

low-fidelity approaches were useful in generating ideas, they were unable to capture the dynamic interactions that a scholar encounters when working with rich collections of cultural resources. Where such a scholar is not particularly technically sophisticated, paper prototypes can be very engaging in encouraging discussions about design. They can help in considering a static presentation of a certain set of results, but can prove difficult in supporting the envisaging of a dynamic interaction. By contrast, high-fidelity prototypes do support consideration of interaction but are slow to develop and thus difficult to iterate within the available schedule, leading to design lock-in.

Desirable visualizations for the Opening History Collection Dashboard were generated through interaction with approximately one hundred library & information science researchers, digital humanities scholars, librarians, archivists and museum technology professionals who attended five different conferences. Two of these design exercises took place took place within l conference exhibit booths. Participants in the sessions were provided with paper templates of visualizations extracted from research papers, known visualization projects and examples from other library, archives and museum interfaces. These cut-out visualizations had small magnets attached so that they could be easily added to or moved around on a light metal board. Although these visualizations represented various other kinds of data, participants were also asked to add annotations that suggested modifications or particular use cases for cultural heritage collection, including drawing desired visualizations on sticky notes.

While some participants were intrigued by more novel visualizations (such as one which showed the physical scale of different sized objects), most indicated a strong preference for visualizations that addressed the traditional questions of who, what, where, and when [21]. Participants were encouraged to do ego-centric design by suggestions like "can you put the items together to make an interface that you personally might find useful?" The separate discussions of what was selected, what was rejected and what was missing were then used to inform a design for less idiosyncratic use. Many participants were familiar with similar kinds of configurable component displays such as in iGoogle and Yahoo! although there is very little like it for online access to digitized collections. Participants noted the desirability of personalization, describing which kinds of elements they would find useful. However this was not just a matter of per-person customization, but also per-use-type. Several people described different activities that they did that would benefit from different combinations of components.

Using input from participants in our demonstrations, we developed an initial prototype that provides a minimal level of functionality for visualizations of Spatial Coverage (Where), Date Items Created (When), Item Types in Collection (What), and a limited list of the top 50 Subjects (What) based on the IMLS DCC collection-level OAI-PMH metadata. Value frequencies were generated using the SIMILE Gadget utility and converted into a comma-separated-value (CSV) tabular format. These individual value/frequencies were then visualized using the Google Visualization API or Many Eyes™ services. The IMA Museum Dashboard, a Drupal template module, was used to stitch together individual visualizations into a complete dashboard . [22]

Figure 1 Opening History Dashboard Prototype

While this prototype provides an overall view of the IMLS DCC aggregation (based on collection-level metadata), it does not yet provide a similar view for each of the collections with item-level metadata aggregated by IMLS DCC. As discussed below, the heterogeneity and quality of item-level metadata posed significant challenges to quickly generating the data necessary to provide accurate visualizations

3.2 Metropolitan Museum of Art

The Metropolitan Museum of Art collections include more than two million works of art from all over the globe spanning five thousand years of history. Information about the collection is presented online in a straightforward manner – faceted by department, searchable by keyword, and sortable by individual fields. The only aggregate collection level information presented is the number of objects in a department. Individual object pages provide an image of the object along with the object description and attendant metadata, but no efforts have been made to further contextualize the object within the collection.

Inspired by other efforts at making cultural heritage information more transparent, the Museum Pipes project started as an examination of the character and quality of the results available through web services retrieved when passing the most minimal of metadata, e.g. artist name, date, or a single keyword [23]. To date, Museum Pipes dashboards have used both Curatorial Department and Museum Library datasets. These early prototypes have been intended exclusively for use by museum professionals. Though subsequent and more thoroughly designed versions of these dashboards could eventually serve the needs of a broader audience, in their current state they are best suited to more specialized tasks. In particular recognizing patterns and exceptions in item level metadata, made more recognizable by their presentation in aggregate, suggesting which collections may require more detailed scrutiny before online publication.

The most robust object level tool from Museum Pipes is an information aggregator that, when provided with a Metropolitan Museum online unique object identifier, scrapes metadata from the Object Record View page on metmuseum.org and passes this to a wide array of third-party information repositories. The current implementation collects press accounts (from NPR, the New York Times, and the Guardian), related images (from Flickr, TinEye, and Google Images), books (from OCLC/WorldCat, Internet Archive, OpenLibrary, and Google Books), additional metadata (tags from OCLC identity records, Wikipedia/DBpedia data, and keywords extracted from the object description by other web

services), along with additional social web content (from delicious, twitter, and YouTube).
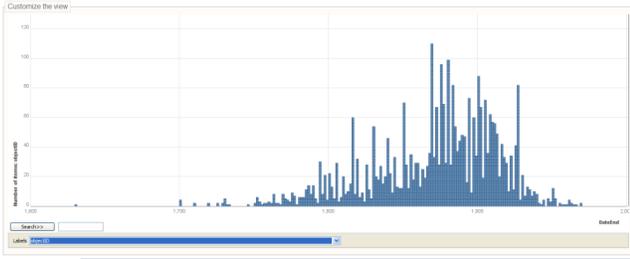
Figure 2 When were works in the American Paintings and Sculpture Collection Made?

4. Discussion

Even within these two case studies, a number of different issues arose – either challenging our ability to develop desired visualizations or raising questions about what kinds of visualizations might be the most useful to target audiences.

4.1 Choosing Visualizations

During our demonstration sessions, we provided users with a variety of different visualization examples, inviting them to think about how such visualizations could be applied to collections information. Currently available web services provide a number of different kinds of visualizations, but often these have been geared towards other kinds of common statistical data – and not the rich textual descriptions commonly found in cultural heritage repositories. This proved challenging on two different levels. Firstly, many of the participants were familiar with text-based collection descriptions and had not previously considered how a visual representation of collection-level information might be used. The challenge here was not only to consider what attributes of collections would be most useful, but how those attributes would be visualized. The IMLS DCC prototype has relatively simple maps, bar charts and slightly more exotic treemap and bubble chart visualizations that were easily grasped by participants.

Additional research is needed to understand how to best provide users with multiple levels of understanding about a collection. Our discussion with participants also suggests that offering dynamic and user-centered ways to view collections. As designers we may not be able to predict how an individual user might want to see subjects across a time scale or the relationship of titles to item types [24].

Furthermore, additional research is needed to understand how these types of visualizations can be blended into traditional search and retrieval interfaces. The prototypes we have been exploring rely on somewhat static and pre-computed visualizations based on established "collection" organizations. However, participants suggested that such visualizations would also prove useful for looking across collections or to understand the contours of a search result set.

4.2 Metadata Quality (Input)

Within the cultural heritage community, issues surrounding metadata quality have been the focus of research for some time. While much of this work has argued for ways to improve search and retrieval services, metadata quality also has a significant impact on the quality and usefulness of visualizations for end users. Developing a visualization based on collection-level metadata created by IMLS DCC was relatively easy because of the metadata's consistency. However we were unable to quickly develop prototypes that worked for any arbitrary OAI-PMH set available through IMLS DCC. While individual OAI-PMH sets could be visualized, internal inconsistencies required significant human intervention to prepare the metadata for visualization. Solutions that worked for one set did not apply to adjacent sets, inhibiting the development of a generalized OAI-PMH visualization service. While the data from the Metropolitan Museum of Art represents a single institution, inconsistencies emerged across data from the nineteen different departments. Within the museum, field names, content and metadata semantics are still vigorously debated from schema development through data entry. The problems with a single institution are compounded across the multiple institutions found in IMLS DCC and maybe further exacerbated by crosswalks from local schema to common OAI-PMH Dublin Core.

A further challenge to dashboard visualizations comes from the services themselves. Because each of the APIs and web services may transport their data in any number of formats, dashboards and aggregators need to be able to translate metadata from its native format into formats required for a particular service and then further package output for use by other visualization components. In the Museum Pipes applications, JSON has been the preferred ingest language because of its nature as a self contained, lightweight interchange format. More expressive or appropriate data standards, e.g. SKOS, have not been very helpful since their implementation still needs to be examined (what fields are used, how they map to other data standards in the same tool) to assure consistent use. RSS has been a consistent choice for output from the Museum Pipes tools. Again, what it loses it expressivity or context specificity it makes up for in reusability of content across contexts – on web pages, in Java applications, or as an easily ingestible format by other services.

5. Future Research

The research that we have undertaken has been ambitious. While information visualizations and dashboards have increasingly entered the mainstream they are still novel products for cultural heritage materials. Our challenge is not only to re-imagine how aggregate information about collections might be presented but also how it might be done differently, for different audiences and different purposes. One approach to this problem could be the kind of in-depth analysis and requirements capture that precedes the development of high-fidelity prototypes (that also come with high costs in both time and funding). But lower-barrier mashup techniques allow us to more rapidly explore the design space than in the past. In essence building visualizations that help us decide what to visualize becomes a viable research strategy. The two cases presented here exploit the possibilities of tightly integrating rapid analysis and rapid design by exploiting the availability of online data and processing resources.

Several important design implications emerge from this initial work:

The initial prototypes developed here will be useful for guiding ongoing discussion about collection dashboards. Early demonstrations that solely relied on paper prototypes left participants struggling with the concept. Later demonstration (with rudimentary mockups) were able to more quickly focus on

the possibilities and problems of visualizing collections information.

Although it is possible to create prototypes with good metadata, inconsistent, incoherent and "unsharable" metadata creates significant challenges to creating useful visualizations.

Even so, good metadata does not guarantee good visualizations. Adherence to a consistent format is not enough to produce a compelling visualization. While an individual repository may internally offer quality metadata, the semantics of metadata across collections/repositories prevents developing aggregation-level visualizations based on item-level metadata. This inhibits the ability of users to usefully make comparisons between and among collections.

Constraints on metadata schema are good for visualization. The more straightforward the schema for item-level descriptions, the more likely it is to be usable across a wide range of tools. This can run counter to the goals of those writing item-level descriptions for subject matter experts, e.g. expressibility, completeness.

Collections of different types of materials (e.g. books in the library, manuscripts in the archive or objects in the museum) may only relate along a single property and have other properties that are unique to the type of material. Additional research is needed to understand how to visualize these unique properties alongside shared properties.

The challenges presented by visualizing heterogeneous metadata offer another possible use-case for visualizations - as a diagnostic tool for metadata creators and their supervisors. Just as financial information dashboards alert traders to fluctuations in the market, integrating a dashboard into metadata creation workflows could assist with quality control efforts. Likewise they may assist service providers who are aggregating OAI-PMH metadata and need to identify how newly harvested sets fit within an existing aggregation.

This research also raised an important distinction. Do these visualizations represent the features of a cultural heritage collection or are they better representations of metadata features? Addressing this question has implications not only for the development of collection dashboards, but also for the properties we choose to include in item-level and collection-level metadata schema and publicly shared metadata.

6. Acknowledgements

7. REFERENCES

[1] Duff, W.M. and Johnson, C.A. Accidentally found on purpose: Information-seeking behaviors of historians in archives. Library Quarterly 72, 4 (2002), 472-496.

[2] Tibbo, H.R. Primarily history: historians and the search for primary source materials. Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, (2002), 1–10. 1.

[3] Lee, H.L. What is a collection? Journal of the American Society for Information Science 51, 12 (2000), 1106–1113.

.

[4] Foulonneau, M., Cole, T., Habing, T.G., and Shreeves, S.L. Using collection descriptions to enhance an aggregation of harvested item-level metadata. Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries., ACM Press (2005), 32-41.

[5] Viegas, F.B., Wattenberg, M., Van Ham, F., Kriss, J., and McKeon, M. Manyeyes: a site for visualization at internet scale. IEEE Transactions on Visualization and Computer Graphics 13, 6 (2007), 1121.

[6] http://vizlab.nytimes.com

[7] Heer, J. and Boyd, D. Vizster: Visualizing online social networks. Proceedings of the 2005 IEEE Symposium on Information Visualization, (2005), 33–40.

[8] http://www.data.gov

[9] Greene, S., Marchionini, G., Plaisant, C., and Shneiderman, B. Previews and overviews in digital libraries: Designing surrogates to support visual information seeking. Journal of the American Society for Information Science 51, 4 (2000), 380–393.

[10] Derthick, M. Exploring Meta-Data Associations with Bungee View. (2007). http://www.cs.cmu.edu/~mad/2007InfoVisContest/BungeeView.pdf

[11] Kramer-Smyth, J., Nishigaki, M., Anglade, T. ArchivesZ: Visualizing Archival Collections. http://archivesz.com/ArchivesZ.pdf

[12] The Visible Archive http://archivesz.com/ArchivesZ.pdf

[13] Few, S. *Information Dashboard Design*. O'Reilly, Cambridge, MA, 2006.

[14] Legrady, George Making Visible the Invisible: Seattle Library Data Flow Visualization. Digital Culture and Heritage. Proceedings of ICHIM05.Paris. (2005).

[15] http://dashboard.imamuseum.org

[16] Stein, R. Transparency and Museums – Walking the Talk. Indianapolis Museum of Art Blog. http://www.imamuseum.org/blog/2009/11/03/transparency-and-museums/

[17] Jones, M.C., Floyd, I. R., Twidale, M. B., Patchworks of Open-Source Software: High-Fidelity Low-cost Prototypes. In St. Amant, K. and Still, B. *The Handbook of Research on Open Source Software*. ISR, Hersey, PA, (2007). 126-140.

[18] IMLS DCC Opening History http://imlsdcc.grainger.uiuc.edu/history/

[19] Palmer, C. L., Knutson, E.M., Twidale, M.B., Zavalina. O. Collection Definition in Federated Digital Resource Development In *Proc ASIST 43*. Austin, TX.

[20] Twidale, M., and Urban, R. J. Usability Analysis of the IMLS Digital Collection Registry, 2005. http://imlsdcc.grainger.uiuc.edu/3YearReport/docs/UsabilityReport1.pdf

[21] Buckland, M., et al., Access to Heritage Resources Using What, Where, When, and Who, in J. Trant and D. Bearman (eds.). *Proceedings, Museums and the Web 2007*.

[22] http://www.richardjurban.net/dashboard/

[23] http://museumpipes.wordpress.com

[24] Card, S. Mackinlay, J. and Shneiderman, B. *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufman (1999).