

Polar Bear Expedition Digital Collections: Enhancing Online Use through Digital Curation

Elizabeth Yakel (yakel@umich.edu), University of Michigan, School of Information, 1085 South University, Ann Arbor, MI 48109-1107 (yakel@umich.edu)

Polly Reynolds, University of Michigan, Bentley Historical Library, 1150 Beal Avenue, Ann Arbor, MI 48109-2113, (pbenes@umich.edu)

Seth E. Shaw (seth.e.shaw@gmail.com), Jeremy York (jjyork@umich.edu), and Magia Krause (mghetu@umich.edu), University of Michigan, School of Information, 1085 South University, Ann Arbor, MI 48109-1107.

Abstract

In this paper, we evaluate the first year (January 2006-February 2007) of operation of the "Polar Bear Expedition Digital Collections" focusing specifically on documenting visitor use of the site, the creation of online communities around records, and the reuse of digital objects. This site is part of a larger research project investigating the ways in which archival finding aids can be re-envisioned and redesigned to increase access to primary sources, particularly through the incorporation and use of social navigation features.

Introduction

Digital curation can be thought of in several different ways: preserving digital content, managing that content over time, and facilitating reuse of digital objects. This last aspect is the least explored and relates more to the definition of a curator in the museum sense; that is, a person who interprets and contextualizes objects for the public. Digital curators put substantial energy into creating standards, developing practices, and heightening awareness of preservation issues; however, the reuse of digital data is particularly problematic because it involves (re)presentation and interpretation. These issues are relevant today when Web 2.0 features, such as social navigation and interaction, raise the specter of shared authority and offer the benefits and challenges of incorporating interaction histories into metadata, thus influencing future public understanding of these objects.

This paper discusses the first year (January 2006-February 2007) of operation of the "Polar Bear Expedition Digital Collections" (<http://polarbears.si.umich.edu>). This site is the first in what will hopefully be a series of experiments by the Finding Aids Next Generation Research Group (FANG) at the University of Michigan School of Information. The FANG team experiments with methods of reimagining finding aids to better support resource discovery, and the use of social navigation to enhance accessibility.

The Polar Bear Expedition Collections are a group of 65 collections of personal papers held at the Bentley Historical Library at the University of Michigan. The "American Intervention in Northern Russia, 1918-1919" (nicknamed the "Polar Bear Expedition") collections document the intervention of the U.S. military into Northern Russia at the end of World War I to fight the Bolsheviks. At that time, military units were geographically organized, so many of the soldiers involved in this event were from Michigan. The Bentley, an archives documenting Michigan history, began collecting these materials in the 1960's and has amassed one of the most comprehensive collections, consisting of diaries, letters, photographs, an oral history, and a motion picture. In 2004, these collections were digitized to increase access and help preserve

the original materials. We selected this group of digital objects for our initial experiment because of the uniqueness of experimenting with entire collections that had been digitally reborn.

We began planning the site in January 2005 and the "Polar Bear Expedition Digital Collections" site went live in January 2006. The site, based on the Everything Development Engine (www.everything2.com), uses a MySQL database as the backbone of the content management system and ImageMagick to render the images. Everything2 was selected because it supported the types of social navigation functions with which we wanted to experiment, specifically commenting, collaborative filtering, and user awareness. In addition to the 12,000+ digital images from the collections, we reused data from other sources including EAD finding aids, MARC records, and a database listing over 6100 soldiers in this campaign.

The "Polar Bear Digital Collections" are also unique in that there is a "real life" community that we have attempted to draw into our online site. This community is primarily composed of descendents of the soldiers. There is also an association which meets each Memorial Day to commemorate the memories of the men involved in the campaign. In addition to this "real life" community, there are numerous historians, World War I aficionados, and others interested in different aspects of the event. Our goal was to create a place where this online community could meet, exchange information, and enhance our online descriptions. In the findings, we discuss the visitors and their uses of the site – both its traditional and social navigation components.

Literature review

Social navigation is defined as an information system which supports collaborative activity (Dourish and Chalmers 1994). Although authors in the archival field have not specifically called for social navigation tools, they have asked for changes in the way we conceptualize finding aids. Duff and Harris argue for shared authority in finding aids to accommodate multiple voices. Light and Hyry promote transparency in the creation of finding aids and the interpretation of archival materials. To date, few archival access systems include these features.

The Science and Technology in the Making (STIM) Project (<http://sloan.stanford.edu/index.htm>), begun in the late 1990's, was the first attempt at creating more interactive archives. This series of 5 websites was aimed at collecting oral histories and recollections rather than representing existing collections. In his final report, Coleman (n.d.) notes "It is difficult to transform communities that have a life outside of the Web into communities that work on the Web unless they believe they are doing real work, ...to be successful ... projects should attempt to integrate the work of the community they wish to address."...'To be 'sticky', sites will need to be self-sustaining and provide an archival presence. By self-sustaining, we mean having the ability to continue to generate interesting and useful content; by archival, we mean having the ability to manage that content for the benefit of its users." This problem of a lack of critical mass or audience was also faced by other non-archival projects. In a 2003 study of online recipes by Svensson and Höök, they note that when too few users are simultaneously logged it is impossible to "see any interesting effects of the real-time presence of others" (201-202).

Currently, archival websites are experimenting with different social navigation features. The Haags Gemeentearchief (<http://www.denhaag.nl/smartsite.html?id=37609>) allows visitors to comment on local photographs. The Everglades Digital Library (<http://cwis.fcla.edu/edl/SPT--Home.php>) lets user's rank digital objects and the Ohio Memory Project provides the ability to create and share online scrapbooks (<http://www.ohiomemory.org/>). While these are all interesting approaches to incorporating social navigation features and shared authority in archival sites, no evaluation of these endeavors has been done. Thus, we have incorporated evaluation components into our research from the beginning.

Methods

The major research questions for the FANG project are whether social navigation functionalities can support resource discovery and enhance accessibility, and if so, which ones. Evaluation mechanisms include web analytics (internal transaction logs as well as Google analytics since August 2006), a survey (March 2006 with another currently in progress), interviews with visitors, and content analysis of visitor contributions. Our goal is to triangulate these data to get a picture of both the quality and quantity of use of the features that support social navigation.

Analytics data from the site itself (the transaction logs and the contributions of visitors) is the best source of information. Response to the initial survey was sparse (6 respondents most likely due to an overly randomized survey pop up on the site) and only three visitors to the Polar Bear Digital Collections site agreed to be interviewed. Therefore, the survey and interviews should be considered anecdotal and illustrative of usage patterns identified elsewhere.

Enhancing Use through Digital Curation

Between 15 August 2006 and 15 February 2007, there were 24,916 visits to the site. Peterson (2005) claims that better metrics on usage are length of time spent on the site and page views per visit. He argues that visitors who spend over 90 seconds on a site are 'real' users and that 'committed' visitors will access between 5 and 10 pages per session. Figure 1 shows time spent per visit; 4,729 visits lasted for over one minute. Figure 2 notes the number of pages viewed per session; we found that 6 or more pages were viewed in 2,943 sessions.

Figure 1: Time Spent (per second) 15 August 2006-15 February 2007

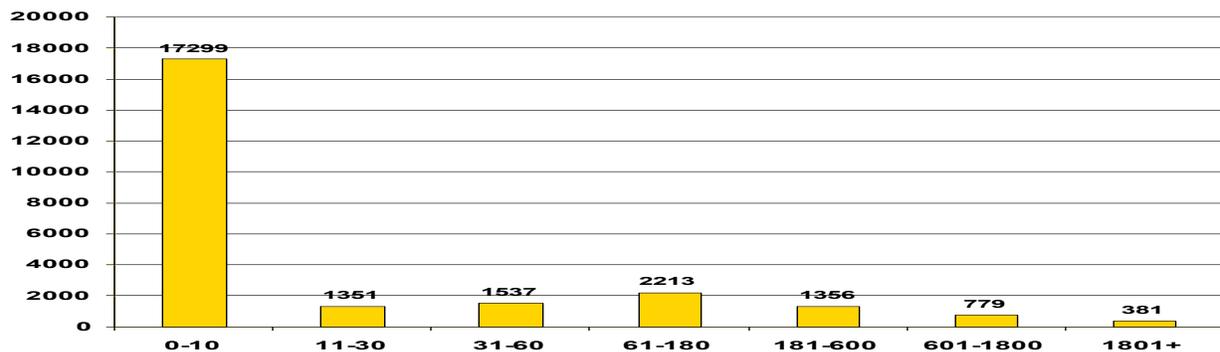
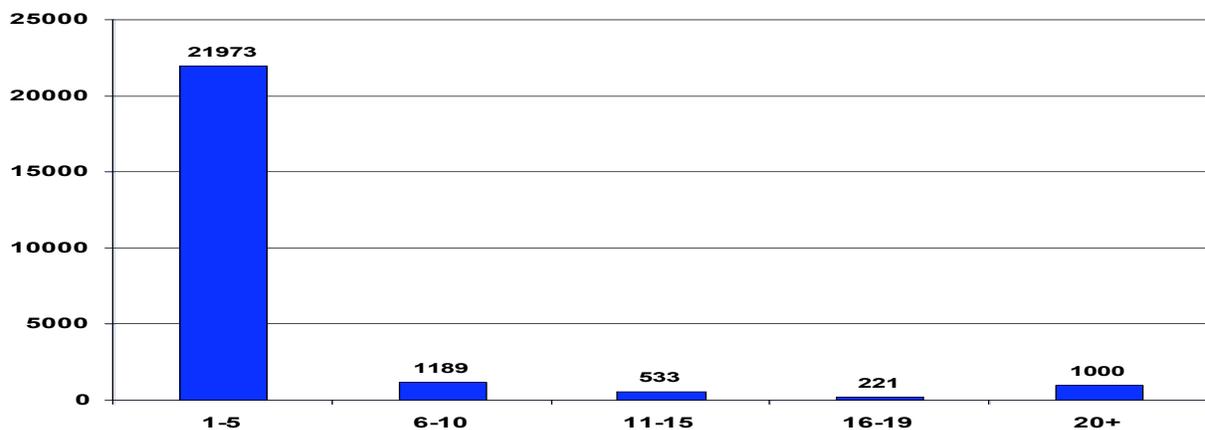


Figure 2: Pages Viewed per Session 15 August 2006-15 February 2007



Visitors reach the Polar Bear Expedition site in a number of different ways. The most common is through Google search (16,223) and the second most common through direct entry of the site's URL. Referrals from other sites have been important as well. Links to the Polar Bear Collections were added on Wikipedia, the Internet Public Library, the Bentley Historical Library website, the Polar Bear Memorial Association website, and MIRLYN, the University of Michigan online bibliographic catalog. Table 1 shows the ranking and frequency of links from these referring sites. Wikipedia and the Bentley were the most fruitful referrals, MIRLYN (MARC records) was the least productive.

Table 1: How Visitors get to the Site (Selected Rankings)

Rank	Source [Medium]	Visits
1	google	16223
2	Direct (polarbears.si.umich.edu)	1861
5	en.wikipedia.org	425
6	bentley.umich.edu	402
10	ipl.org	267
26	pbma.grobbel.org	39
43	mirlyn.lib.umich.edu	13

The web enables use from around the globe. Table 2 outlines the top 4 countries from which visitors hail. Most are from the United States, but international interest is evident with a number of visitors from Arkhangel'sk (Archangel), the area in Russia where the intervention took place. Within the U.S., 2620 (almost 16%) of the visitors are from Michigan; they also view double the number of pages than others in the U.S. when on the site.

Table 2: Where Visitors Come From: Top 4 Rankings

Rank	Country/State or City	Number of Visits	Page Views per Visit
1	United States / Michigan	16527/2620	4.88 / 9.7
2	Canada	1855	3.42
3	United Kingdom	1064	3.11
4	Russian Federation / Arkhangel'sk	939/97	11.18 / 38.14

All visitors to the Polar Bear Expedition site are invited to register. Registration has several benefits: the ability to contribute comments, bookmark information or images, see simultaneous visitors, and provide a user profile. As of February 2007, 221 individuals formally registered. Of those registrants, 19 (9%) completed user profiles. Analysis of these profiles shows that most (13) have a family connection to the event. Of the remaining 6 who have submitted online profiles, 4 indicate an interest in the history of the event or the era. From the registrants, we see that the preexisting, but not necessarily connected, Polar Bear Expedition community is slowly building an online community around the site. While the user profiles detail an individual's particular interest in the Polar Bear collections, users also contribute other information as well. Such contributions can be divided into three categories: additional information about an individual or updates on the family of the individual, questions about an individual or searching for information, and information sharing (in several cases adding a URL with a link to a photograph or diary. A typical user profile reads: "My Great Uncle ... was killed in action February 1919. ANY information regarding his death or his service with the Polar Bears would be greatly appreciated!!!!!! I have a book printed after the great war with listings and photos of soldiers who were killed in WWI including the AEF in Russia. The Problem is the photos are listed by state and are not in Alphabetical order...If you have a relative who was killed during that time I may have a photo of them in this two volume set. my email is ... if you wish to contact

me” (User profile node #7297). Most of the people in the user profiles are identifiable and several others, in addition to the author of the previous quotation, provide contact information. We surmise from this behavior that the online community puts some degree of trust in the site.

The original Polar Bear collections at the Bentley were frequently requested. We were interested to see whether digitization, which would dramatically increase exposure to the collections, had any effect on use of the originals. We found that digitization has dramatically increased exposure to the collections and usage in the Bentley has declined. We examined call slips for the physical collections at the Bentley from July – December 1999, a typical period before digitization and from July – December 2006, post-digitization. Table 3 shows that virtual use of the collections exceeds previous and present use of the collections in the Bentley.

Table 3: Use of the Polar Bear Collections Pre and Post Digitization

	At the Bentley	Online
1999 (July – December)	51	n/a
2006 (July – December)	5	17,523 visits

Enhancing accessibility of the Polar Bear Expedition site has been done through design as well as the addition of social navigation features. In terms of the design, the FANG research team strove to make the site intuitive and one that a visitor could easily “walk up and use”. Early on in the site development, we decided to create a rich browsing structure containing 7 categories: collections, individuals, military units, geographic locations, subject, media type, and organizations. We did this using enhanced mark-up of EAD finding aids and reconciliation of similar MARC subject headings. As a result of our efforts, researchers have much easier access to the site’s collections than that provided by traditional EAD browsing systems, which tend to use browsing by creator or collection title only.

Table 4: Browse vs. Search

Page Title	Unique Views	Page Views
Homepage (Welcome)	5411	7457
browse by: geographic location	1859	2911
new advanced <i>search</i>	1822	4162
browse by: individual name	1566	3614
browse by: collection	1448	2578
browse by: media type : Photographs.	1275	2291
Polar Bear History	1126	1344
browse by: military unit	830	1579
United States Army Signal Corps photograph collection	692	1600
Frank J. McGrath photograph album.	592	740

The creation of this rich browsing structure is quite successful. The web analytics demonstrate an overwhelming preference for browse over search. Browsing by geographic locations and individual names lead other browsing options in popularity. Table 4 shows the top ten pages viewed on the site, five of which are browsing lists. The browsing categories of media type and subject, while not on this list, are in the top 20 most frequently consulted pages. As one interviewee noted, “My preference has basically been to browse by just going through type of subset and then either alphabetically or whatever then go down, scroll down the list until I find what it is I’m looking for” (Interview 1, section 19). This preference for browse over search adds

evidence to recent research calling into question automatic preferences for search when appropriate browsing structures are available (Katz and Byrne 2003; Olston and Chi 2003). Bookmarking, a design feature we experimented to enhance basic access, was less successful. Although the interviewees and survey respondents liked the bookmarks, only 8 people used the bookmarking feature in the first year. Visitors created 35 bookmarks, ranging from one person with 19 to 4 visitors with 1 bookmark each (the mode). When we implemented the bookmarks, we thought they would reduce the drain on short term memory and enable customization of the site; however, unlike del.icio.us.com, bookmarks on the Polar Bear Expedition site are not sharable and visitors cannot use them to create communities around common interests.

The Polar Bear Expedition site utilizes several features that foster social interaction: user profiles, visitor awareness, link paths, and comments. These features encompass both direct and indirect interaction (Höök, David Benyon, and Alan Munro 1999). Direct interaction involves explicit action on the part of the visitor, and occurs in such situations as asynchronous responses to comments and synchronous online chat. During direct interaction, people are aware of one another. Indirect interaction, on the other hand, guides people less explicitly through information left as a byproduct of others' activities (Dieberger 1999). Indirect interaction is also referred to as collaborative filtering because it aggregates information about previous users (collaborating) to make automatic predictions (filtering) about what future site visitors will want (Goldberg et al. 1992). Amazon.com's book purchasing recommendations are a well-known example of indirect interaction. On the Polar Bear Expedition site the user profiles and comments are direct social interaction mechanisms; link paths are indirect. User profiles have been discussed previously so we focus on the 'link paths' and comments in this section.

Link paths, originally developed in Everything2 as "softlinks", are one of the collaborative filtering, or recommender, mechanisms used on our site. They provide feedback to visitors concerning the navigational paths previously taken to reach a particular item or collection. As more people use the site, more link paths are aggregated and the overall filtering mechanism (displaying what visitors who viewed this item also viewed) becomes more refined. Over time, we hope that link paths will help to uncover new and unexpected connections between subjects and collections. Paper finding aids have physical signs of use – dog-eared pages and annotations – we hope that the link paths will serve as a virtual equivalent.

Link paths were somewhat problematic for us in the beginning. Survey respondents and interviewees did not initially comprehend the list of links appearing at the bottom of each page titled "link paths". Therefore, we made several adjustments. First, we added a tag line to the link paths similar to that used by Amazon.com: "Researchers who viewed this page also viewed". We also adjusted the link paths' algorithm. At first we set the bar for populating the link paths too high and as a result, they were not being populated quickly. When they were, the homepage and help pages appeared at the very top. While the homepage and help pages were common pathways from many pages, they did not provide us with the kind of relationships and links between collections we wanted. As a result, we eliminated the help and the home pages from the algorithm. Additionally, for privacy reasons, we removed user profiles from the link paths as it might identify a particular user's paths and we wanted the link paths to be anonymous aggregations of users' movement. By redesigning the link paths algorithm to display only collections, digital images, or biographical information from the soldiers' database, we think this feature will better support the relationships and connections we are trying to foster.

The Polar Bear Expedition Digital Collections incorporates a commenting feature which allows users to leave lengthy descriptions and encourages asynchronous interaction among visitors. Comments also become part of the overall system and are searchable along with other text on

the site. The comment feature has been the most utilized of the social interaction mechanisms on the site. Twenty-nine visitors posted 62 comments between January 2006 and January 2007. Interestingly, of those 29 people, 7 (25%) also created user profiles. Thus, there is some overlap between individuals who use these features. This group is probably the core community interested in the site; however, it is still relatively small and has not achieved critical mass.

Comments can be grouped into 3 categories: information sharing, question asking, and donation inquiries. Information sharing can also be divided into two types: descriptive data and error correction. The tenor of these is similar to that in the user profiles: visitors post links to related online information or provide information about soldiers or images. For example one expert in subchaser boats identified photographs and provided detailed information: "Eagle 2 is one of the U.S. Navy Eagle boats built by Ford Motor Company. Eagles 1, 2 and 3 were sent to northern Russia in 1919. See, for example: <http://subchaser.org/set-russia-05>".

Other comments identify potential errors or omissions. A typical correction message reads: "Incidentally Roy passed away January 17, 1955 following a battle with pancreatic cancer and is buried in Downing Cemetery, Deckerville, Mi." (Comment on Roy W. Randall 6/26/2006). We encourage this type of interaction and had anticipated that this would happen. We ask visitors for evidence, such as a death certificate or discharge papers. In most cases, we have updated our information after reviewing the documentation. If a user is unable to provide documentation, we do not change the site, but keep the comment to allow for multiple voices.

Visitors also ask questions, some diffuse and some specific, in the comment feature. "I was always 'told to remember that grandpa was a Polar Bear' ... I am very interested in the history involved. Anyone with information on Frank McGregor included in their families things I would greatly appreciate. Thank you. This is a great site" (Comment on Frank McGregor, 7/18/2006).

Donation inquiries are the third type of comments. We were unprepared for these questions from potential donors which included requests to post their digital pictures on our site. "The Only thing I would have to donate to your collection would be anything digital I have of Oliver's [O.A. Mowat]... If you would like me to send digital pictures of things I have of Oliver I would be happy to" (Comment on O.A. Mowat 3/21/2006). Donations posed a problem since we are a research project and not a collecting repository. We refer all donation questions to the Bentley Library and they can decide how items fit into their online and/or physical collections.

In addition to the user profiles and the comment features, visitors can contact "the Archivist" through a 'Contact Us' area. These messages are also public and searchable. In reality, visitors use all three of these areas to send messages to "the Archivist". In fact, most of the dialog on the site is between visitors and "the Archivist". There has been some interaction between the researchers themselves, particularly among individuals who know each other from the 'real world' Polar Bear community. In one exchange, one visitor suggested additional websites and other historical resources to another researcher in search of information on her ancestor. We hope to encourage and see more user-to-user interaction in the future. In particular one member of this community responds to questions and provides information to others.

Discussion and Conclusions

We have learned several things from this experiment. First, browsing structures are useful/powerful because visitors can see what is there and do not have to pull search terms out of their heads. Archivists should not assume that users prefer search over browse. Second, thinking more broadly about how to increase access is important. Posting information about primary sources in popular sites, such as Wikipedia and the Internet Public Library, does lead to

referrals. In our case, this led to many more than from MARC records. Finally, creating a robust online community around records is difficult. While an entire repository may have a substantial following, smaller groups of records may not have sufficient audiences to create dynamic online communities. We still believe that social interaction holds great promise for increasing access to and improving descriptions of archival materials; however, we found that some features (e.g., comments) appear to work better than others. More (long term) experiments and projects that explore and evaluate other mechanisms for social interaction, such as annotation, ranking, and tagging are desperately needed. Each of the social navigation features described in this article establishes a set of affordances and precludes others. It will not be until we have experimented and studied the multiple options that exist in the digital environment that we will be able to best represent and interpret primary sources to all of the potential audiences in virtual space.

Acknowledgements

The authors thank the staff of the Bentley Historical Library for allowing us to use the Polar Bear Expedition Collections. We also acknowledge contributions of team members Christie Peterson, Dharma Akmon, Andrew Bangert, Ricah Marquez, and James Sweeney, our first programmer.

References

- J. Coleman (n.d.) Final Report for the STIM Core Site. URL:
<http://www.stanford.edu/dept/HPS/sloanconference/CoreSiteReport.html>
- A. Dieberger, "Social Connotations of Space in the Design for Virtual Communities and Social Navigation," in *Social Navigation of Information Space* Alan Munro et al. eds. Computer Supported Cooperative Work. London: Springer, 1999.
- P. Dourish & M. Chalmers, (1994). Running Out of Space: Models of Information Navigation." In *Proceedings of Human Computer Interaction*. (HCI'94) 1994.
- W. Duff & V. Harris. (2002) Stories and Names: Archival Description as Narrating Records and Constructing Meanings. *Archival Science* 2/3-4: 263-285.
- D.Goldberg, D.Nichols, B.M. Oki and D. Terry. (1992). Using collaborative filtering to weave an information Tapestry. *Communications of the ACM* 35/12: 61-70.
- K. Höök, D. Benyon, & A. Munro. (1999). Editors' Introduction: Footprints in the Snow. In *Designing Information Spaces*. K. Höök, D. Benyon, A. Munro eds. London: Springer: 1-16.
- M.A. Katz & M.D. Byrne. (2003) Effects of scent and breadth on use of site-specific search on e-commerce Web sites. *ACM Transactions on Computer-Human Interaction* 10/3:198 - 220
- M. Light & T. Hyry. (2002). Colophons and Annotations: New Directions for the Finding Aid. *American Archivist*. 65 (Fall/Winter): 216-230.
- M. Svensson & K. Höök. (2003). Social Navigation of Food Recipes: Designing Kalas," in *Designing Information Spaces*. K. Höök, D. Benyon, A. Munro eds. London: Springer: 201-222.
- C. Olston, & E. Chi, (2003). ScentTrails: Integrating browsing and searching on the Web. *ACM Transactions on Computer-Human Interaction* (TOCHI) 10/3: 177 – 197.
- E.T. Peterson. (2005). *Web Site Measurement Hacks*. Publisher: O'Reilly.