

2011

# Researchers at Work: Assessing Needs for Content and Presentation of Archival Materials

Janet Hauck

Whitworth University, [jhauck@whitworth.edu](mailto:jhauck@whitworth.edu)

Follow this and additional works at: <http://digitalcommons.whitworth.edu/libraryfaculty>



Part of the [Archival Science Commons](#)

---

## Recommended Citation

Hauck, J., Allison-Bunnell, J., & Yakel, E. (2011). Researchers at Work: Assessing Needs for Content and Presentation of Archival Materials. *Journal of Archival Organization*. 9 (2), 67-104.

This Article is brought to you for free and open access by the Library at Whitworth University. It has been accepted for inclusion in Library Faculty Scholarship by an authorized administrator of Whitworth University.

## **Researchers at Work: Assessing Needs for Content and Presentation of Archival Materials**

Jodi Allison-Bunnell, Elizabeth Yakel, and Janet Hauck  
2010 October 29

### **Addresses of authors:**

Jodi Allison-Bunnell  
Program Manager, Northwest Digital Archives  
Orbis Cascade Alliance  
418 Woodford  
Missoula, MT 59801  
[jodiab@uoregon.edu](mailto:jodiab@uoregon.edu)

Elizabeth Yakel  
Associate Professor  
University of Michigan  
School of Information  
1085 South University  
Ann Arbor, MI 48109-1107  
[yakel@umich.edu](mailto:yakel@umich.edu)

Janet Hauck  
University Archivist  
Whitworth University  
300 W. Hawthorne Road  
Spokane, WA 99251  
[jhauck@whitworth.edu](mailto:jhauck@whitworth.edu)

### **Biographical statements of authors:**

Jodi Allison-Bunnell is the Program Manager for the Northwest Digital Archives (NWDA) program at the Orbis Cascade Alliance. NWDA offers enhanced access to over 5,700 archival and manuscript collections in Alaska, Idaho, Montana, Oregon and Washington through a union database of Encoded Archival Description (EAD) finding aids. Her presentations and publications include those on consortium planning and sustainability, usability and user studies, archives advocacy, and copyright issues. She holds an MA and an MLS from the University of Maryland at College Park and a BA *summa cum laude* from Whitman College. She is an active member of the Society of American Archivists and Northwest Archivists.

Elizabeth Yakel is an associate professor at the University of Michigan School of Information where she teaches in the Archives and Records Management and Preservation of Information specializations. Her research focuses on archival user needs and access to primary sources. Yakel has received funding from the Andrew W. Mellon Foundation, the National Historical Publications and Records Commission, and the OCLC Online Computer Library Center. She publishes on archival representation, access, and use in such publications as *American Archivist*, *Archivaria*, and *Archival Science*. She has been active in the Society of American Archivists and was elected a Fellow in 1999.

Janet Hauck is University Archivist and Adjunct Instructor of History at Whitworth University in Spokane, WA. She has published on the Greene and Meissner method of archival processing and presented papers on archival instruction methods and religious history in the Northwest. She holds an MLS from the University of Washington, along with certification from the Academy of Certified Archivists. She teaches public history courses at Whitworth and is an active member of the Society of American Archivists and Northwest Archivists. She received a Faculty Research and Development Grant from Whitworth in 2008 that supported a portion of this study.

## **Abstract**

In the past, programs that present digitized archival materials were often created with limited knowledge of their audiences' needs and greater focus on the materials. Organizations must ask whether digital programs are sufficiently effective to merit financial support. As part of the planning process for a digital program at the Orbis Cascade Alliance, the Northwest Digital Archives (NWDA) consortium conducted a study of its core researchers' needs for the selection and presentation of archival materials online. With the assistance of NWDA members, nineteen subjects were recruited for hour-long interviews. Although the number of subjects meant that the conclusions should be regarded as preliminary rather than definitive, we were able to draw conclusions about these researchers' needs and desires that will shape the development of the Alliance's program.

## **Introduction**

Since the early 1990s, , archives and special collections have created digital surrogates of their analog collections and made them available for their researchers, first on CD-ROMs, then on the Web. These first projects and their successors have been driven by the usual dual concerns of archival institutions: to make unique collections more broadly available (access) and to protect fragile collections by having researchers use digital surrogates (preservation). These efforts also have often been driven by the materials and the archives that hold them and by a “build it and they will come” mentality that suggests that the exposure of materials will naturally build broad audiences for collections. Unlike many portions of the museum and library worlds that have embraced evaluation and user studies,<sup>1</sup> archival collection management systems, websites, and similar tools have been built largely based on the perceptions that archival professionals have about user needs. Now, the days of digitizing special collections materials because someone, somewhere, might be interested in looking at them are long past. Instead, reduced resources and the real need to assess the effectiveness of current programs demand that we more closely study audience needs and preferences and first answer the question, “Why digitize, and for whom?”

The Northwest Digital Archives (NWDA) program at the Orbis Cascade Alliance has undertaken the Researcher Needs Study to shape its future programs to present and integrate digitized archival collections and metadata. NWDA, a consortium of thirty archival institutions located around the Northwest, was funded by the National Endowment for the Humanities and the National Historical Publications and Records Commission from 2002 to 2007 as an EAD finding aid project and became a program of the Orbis Cascade Alliance in July 2007. The Alliance is a

---

<sup>1</sup> See, for instance, the webliography that the Institute for Museum and Library services refers its applicants to at <http://www.ims.gov/applicants/learning.shtm>; its titles were primarily created for museums and libraries.

consortium of thirty-six academic libraries in Oregon and Washington that offers collaborative library services in these and surrounding states; those services include a union catalog, courier service, cooperative purchasing of resources, and digital services including NWDA.<sup>2</sup> NWDA provides enhanced access to archival and manuscript collections in Idaho, Montana, Oregon, Alaska, and Washington through a union database of EAD finding aids.

A prime driver of the NWDA-Alliance merger was the mutual desire to create a digital content program to present locally-held unique materials, including archives and special collections, online. Starting immediately after the merger and concluding in 2009, the NWDA Program Manager conducted a needs assessment and planning process for a larger digital program at the Alliance. The aim of the Alliance staff and NWDA leadership was to create a program that serves documented needs of real people who are priorities for service from NWDA and Alliance institutions. A program that serves documented needs in a proven way is one more likely to inspire long-term support by its member institutions.

In pursuit of these goals, NWDA completed three initial studies:

- Survey of Digitizing Initiatives (October-November 2007): A survey of all forty-eight Alliance and NWDA member institutions to assess their levels of activity in any of fourteen different types of digital projects or programs, and their overall desires for the Alliance's direction in this area (NWDA, 2007). This study revealed that of the many projects and programs to present digitized archival content, few were sustainably funded, and hosting institutions had many questions about effectiveness and audiences served.

---

<sup>2</sup> For more information about the Orbis Cascade Alliance, see <http://orbiscascade.org/>.

- Researcher Type Survey (Winter 2007-2008): A survey of NWDA members only, asking what types of researchers they have served in their archives in the last year, and who they consider their priority types for services. These categories determined the types of researchers the program recruited as subjects for the fourth and final study, the Researcher Needs Study (NWDA, 2008 January).
- Institutional Needs Survey (July-October 2008): A four-part survey of all Alliance and NWDA member institutions to assess their level of interest in different types of programs; needs for best practices and guidelines; training needs; and interest in scanning and reformatting services (NWDA, 2008 October).

The study that is the subject of this article, the Researcher Needs Study, was the fourth and last part of this two-year needs assessment. With the assistance of NWDA members, this study recruited administrators, genealogists, college/university faculty, amateurs/avocational researchers,<sup>3</sup> and college/university students for one-hour telephone interviews with webcasting. Each interview focused on six pairs of national, regional, or institutional sites that present archives and special collections materials. We completed a total of nineteen interviews. The interviews, and subsequent analysis of those interviews, sought to answer research questions in three areas: current use of sites, metadata, and credibility. The goal was to inform the future

---

<sup>3</sup> We use the term “amateurs” as defined by Robert Stebbins: “*Amateurs* are found in art, science, sport, and entertainment, where they are inevitably linked, one way or another, with professional counterparts who coalesce, along with the public whom the two groups share, into a three-way system of relations and relationships. By contrast *hobbyists* lack the professional alter ego of amateurs, though they sometimes have commercial equivalents and often have small publics who take an interest in what they do. The professionals are identified and defined in (economic rather than sociological) terms that relate well to amateurs and hobbyists, namely, as workers who are dependent on the income from an activity that other people pursue with little or no remuneration as leisure” (see Stebbins, 2006, pp. 6-8).

design of a program to present digitized content and to contribute to the profession's knowledge.

Specific research questions in these areas were:

1) Current use of sites

- How are these researchers currently using online archival materials, for what types of work, and how would they like to use them?
- What (design, functionalities, etc.) facilitates their work processes?

2) Metadata

- What elements do researchers most desire in descriptions of digitized materials?
- What types of metadata enable resource discovery and selection?

3) Credibility

- How do researchers assess credibility?
- Do researchers see Web 2.0 functionalities as affecting authority in archival descriptions?

Since the conclusion of NWDA's surveys and study, the findings have shaped the development of digital programs at the Orbis Cascade Alliance. Among other initiatives, and consistent with the findings of this study, the program will move forward with development of a cross-search utility that will contextualize digitized collection materials by creating a new presentation that draws together the digitized object and item data with collection-level metadata from the EAD finding aids at an appropriate level of hierarchy.<sup>4</sup> Also consistent with the findings of this study, this utility will also rely largely on search engines for search and exposure rather than focusing resources on development of an elaborate search interface or destination portal.

---

<sup>4</sup> For a textual and visual representation of the cross-search utility, see [http://orbiscascade.org/index/cms-filessystem-action/nwda/files/cross-search\\_utility\\_mockup\\_20100924.pdf](http://orbiscascade.org/index/cms-filessystem-action/nwda/files/cross-search_utility_mockup_20100924.pdf) and [http://orbiscascade.org/index/cms-filessystem-action/nwda/files/cross-search\\_bullet\\_list\\_20100526.pdf](http://orbiscascade.org/index/cms-filessystem-action/nwda/files/cross-search_bullet_list_20100526.pdf).

## **Literature Review**

Since the early 1990s, the amount of data about archival collections as well as the collections themselves has increased through the Internet. Archivists have been using the web consistently since that time, largely to publish surrogates, i.e. finding aids for collections. Only recently has a critical mass of digital images and entire archival collections become available for study and leisure. Still, relatively little is known about what audiences want in terms of these descriptions or entire collections. This literature review will briefly outline some of the studies of online finding aids and digital images and collections.

There has been relatively little experimentation with the form of online finding aids either in terms of http or EAD. Most online archival description looks much like its paper counterpart. The development of EAD entailed considerable analysis of finding aid structural elements from diverse repositories (Pitti, 1997); however, there was no user input into the development of the standard or the early finding aids utilizing EAD that were published on the web. As the standard began to be adopted, archivists themselves found that adaptation of the display was necessary, but early changes were based on staff perceptions of users, rather than the user themselves (Meissner, 1997).

In the last ten years there have been a number of studies of online finding aids. Two studies performed content analyses on online finding aid systems. Kim (2004) noted large differences in displays, such as inconsistency in the use of data elements, labeling terminology, browsing attributes and limited navigational aids and search functions in many systems. Zhou (2006) focused on the search features and found these to be variable and poor. She also identified

retrieval inconsistencies when searching for the same finding aid in cooperative EAD databases and at the home institution.

These issues as well as others have emerged in the few actual usability tests that have been conducted thus far on online finding aids. Hutchinson's (1997) retrieval study used typical questions nominated by archivists in two finding aid systems under four conditions: 1) searching entire finding aids; 2) searching introductory material to finding aids; 3) searching introductory material to finding aids enhanced by controlled vocabulary terms; and 4) searching collection-level catalog records. He found differences among the treatments in the proportion of relevant documents returned by searches and argued that there is a real value in a field-delimited mark-up of finding aids to facilitate context searching. Hutchinson also found that while a search of the entire finding aid improved recall, precision was radically decreased and the ability to search different sections of the finding aid was critical for precision. Czeck's (1998) examination of subject terms in archival MARC records and finding aids confirms Hutchinson's conclusions about the importance of full text searching in finding aid sections, including the scope and content and bio/administrative history fields.

Studies involving users of online finding aid systems confirm and extend these findings.

Although there have now been a number of these, their diverse methodologies and findings do not allow for generalization. The earliest study of online finding aid systems was Altman and Nemmers' (2001) focus group, which found both usability and archival terminology issues in online finding aids. Hamburger (2004) surveyed archives users to examine their methods of

resource discovery. She concluded that most online search strategies relied on proper names and that users were often dissatisfied with the results.

There have been four published usability studies of online finding aids. The earliest found that users often got lost navigating the hierarchical structure of finding aids and had persistent problems moving between the left and right frames. The use of archival terminology, including search operators, was also confusing (Yakel 2004). Prom (2004) identified differences between novices and experts in his study of multiple finding aid systems. Although all users performed better on the simpler interfaces, novices had trouble with archival terminology and organization throughout the experiment. Subjects also found the desired information more often in the interfaces that provided a clear browsable hierarchy and preferred those with greater navigability, even if this was through the browser's "find" function. Scheir's (2005) study of novices focused on four issues: terminology, navigation, display, and structure, employing known item searches to tease out these concepts. She found that site structure often assumed knowledge of archival practice and principles but that her subjects were able to traverse the learning curve during the study. Over the course of the experiment, subjects developed more efficient searches and felt greater ease and confidence with the system. Still, archival terminology and the hierarchical navigation required to use online finding aids posed difficulties for these users. This study also found that users preferred fewer large text blocks and wanted summaries of fields with large amounts of text. Howard's 2006 thesis studied the placement of the navigation / container information on the left versus the right sides of the screen. She concluded that, "the fact that most participants [19 out of 22] answered all the questions correctly suggests that the placement of the container information did not have an overwhelming effect on

their ability to complete the tasks” (p. 18). However, subjects answered the questions more quickly with a left-side navigation bar and their exit interviews confirmed a preference for this configuration.

In addition to the published usability studies, there have been a number of unpublished studies, principally by major consortia that aggregate online finding aids: the Online Archive of California and the Northwest Digital Archives. The Online Archive of California has done four rounds of usability testing from 2001 through 2009. As a result, the OAC has been able to improve search functions and display. The latest round of usability testing led to an entire redesign of the interface, in particular the display of the finding aids, which was publicly released in June 2009.<sup>5</sup>

The Northwest Digital Archives has also done several rounds of usability testing. This has led to findings concerning level of detail; participants wanted more image content and less text. This is somewhat problematic because they also desired sufficient detail “to get started on their research, which meant enough to see if the collection is likely to contain relevant information and details about what is in each box. However, while detail was expected, many still said they would rather not have to read long blocks of text” (NWDA Test 4, pg. 1). In a later usability test that focused on the search interface, they detected problems in the search interface because subjects saw the browse options as search limiters (NWDA Test 5).

---

<sup>5</sup> See <http://www.oac.cdlib.org/>.

This body of EAD research has three themes: confusion caused by the archival terminology and practice embedded in online archival finding aids, the difficulty of selecting search terms, and problems with navigating through finding aids, particularly large ones. Each of these issues is considered in the present Researcher Needs Study. However, we take a broader approach and examine search processes in the context of systems, rather than analyzing the systems per se, as in a traditional usability study.

The studies of digital images have also been spotty and diffuse. Few of the usability tests have been done with habitual or established users and many concentrate on some specific aspect of the site rather than the entire interface or comprehensive functionalities. One of the biggest foci in studying digital image collections has been metadata. The first study of this type focused on the Library of Congress' American Memory project (Library of Congress, 1991-1993). The biggest surprise of this report was the degree to which K-12 teachers and students were relying on the site and the relative lack of use by LOC's traditional users. More recently, Choi and Rasmussen studied user queries to American Memory. Most users searched for names, things, events, place names or time periods. Historians in the study primarily liked subject and format terms. Minnesota's Foundations Project did usability testing on its interface and made changes to the way its metadata was displayed (Foundations Project, 2000; Quam, 2001). The Cornell University MetaTest project used eye tracking to determine the importance of metadata and how well it was used (Liddy et al., 2002). These different methods and approaches have all pointed out the difficulty in using descriptive metadata to accurately and completely describe digital images.

## **Design and Methodology**

The NWDA Researcher Needs Study employed quasi-experimental methods to address the research questions discussed above. Subjects responded to a controlled script and were asked standardized questions about a series of archival websites. The details of this design follow.

### ***Subjects and Recruiting***

The subject population was determined by the winter 2008 Researcher Type Survey of NWDA members. That survey identified both the major types of researchers that used members' archives and special collections, and the types of researchers considered to be their priority groups for service from their archives and special collections programs. As a result, the five types of researchers targeted in the present study are:

- Staff/administration
- Students
- Faculty
- Serious amateur historians/avocational researchers
  
- Genealogists/family historians

In order to recruit subjects for the interviews, the thirty NWDA member institutions were divided into two groups. One of us communicated with all thirty institutions and urged them to participate actively in the recruiting process in summer 2008. Fifteen institutions distributed a request to five of their most recent in-person researchers in archives and special collections to go to a short online survey to indicate their willingness to participate in the study. Fifteen others distributed the same request to five of the most recent researchers who used collections remotely (e.g., by telephone or through email reference). The solicitation contained a summary of the

study, testing incentives, and time involved. Potential subjects saw a check-box where they could indicate their willingness to participate, characterize what type of researchers they were, and provide their contact information (name, email, and telephone number only). They also had the option to refuse participation. With complete participation of the thirty NWDA member institutions, this process would have yielded a sample of 150 possible subjects.

In reality, and for a range of reasons, this method yielded twenty-nine potential subjects and in the end nineteen actual subjects participated. Four others were willing to participate but were unable to be scheduled or did not show up to scheduled interviews. Experienced researchers with established research skills in using archives and special collections materials made up a majority of the sample. This was not surprising given that our recruitment strategy targeted persons who had recently used collections. The size of this population certainly affected the findings of the study, and its conclusions must be regarded as preliminary rather than definitive. Breadth was deliberately chosen over depth in all aspects of our approach (number of subjects, number of sites viewed with each subject). However, the approach still met the sampling objective of groups identified in the Researcher Type Survey, and it is reasonable to generalize about the needs of these institutions' users from our conclusions.

Five types of researchers were included in this study: administrators, genealogists, college/university faculty, serious amateurs/avocational users, and college/university students. Initially we planned to recruit alumni of the NWDA academic institutions, but this proved impractical, since everyone involved in this study was also an alumnus of at least one academic institution, and his/her primary research needs did not relate closely to that status.

Unfortunately, we were not able to recruit as many genealogists as we would have liked.. Table 1 shows the distribution of subjects by type.

**Table 1: Researcher Types Included in Study**

<b>Type of Researcher</b>	<b>Number</b>
Administrator / staff	8
Genealogist	1
Professor	2
Student	4
Serious Amateur/Avocational	4
<b>Total</b>	<b>19</b>

Subjects received \$20 Amazon.com gift certificates in appreciation of their time.

Most subjects did research online in archival materials for academic work, non-academic work (administrative, particularly buildings and facilities), and avocational work. However, subjects had little or no experience using the sites included in the study. Even those who had used one of the sites before may not have used it to locate archival materials. Subjects who had used any of the sites were most likely to have used WorldCat or Flickr (all twelve sites are discussed below). Extensive users (three subjects) of WorldCat were most likely to have used it for interlibrary loans of books and journals, to determine the extent of information available on a subject, or to find secondary sources. Moderate users (nine subjects) were most likely to have used WorldCat for the same purposes. Subjects had used Flickr to post family or organizational photographs rather than to conduct any type of research. As mentioned, the vast majority of the subjects had never used the sites included in the study.

### *Sites Used in the Study*

We included twelve sites in the testing. These sites were paired to enable subjects to compare and contrast the sites, creating a total of six pairs. These sites exemplified various characteristics to which we sought users' reactions. What follows is a discussion of the sites and the reasons for selecting the pairs.

#### Multi-institutional, format-integrated sites:

- WorldCat (<http://www.worldcat.org/>), with content from the Montana Historical Society on the Montana Memory site
- Flickr (<http://flickr.com/>), with content from the Library of Congress

In the case of WorldCat and Flickr, archival materials (in particular photographs described at the item level) from a variety of institutions (e.g., libraries, archives, museums) are grouped into large databases. The focus of this test was accidental discovery of archival materials among non-archival materials as well as the potential for social computing.

#### Contrasting presentations of the same archival material for a geographic region:

- Online Archive of California (OAC) (<http://www.oac.cdlib.org/search.image.html>)<sup>6</sup>
- Calisphere (<http://www.calisphere.universityofcalifornia.edu/>)

The Online Archive of California (OAC) and Calisphere consist of materials that are pulled together from hundreds of institutions in the same geographical region. The same archival materials are presented in two different ways. In the OAC, archival materials are linked to

---

<sup>6</sup> This study was conducted before OAC's redesign was released in June 2009.

finding aids and the site is designed for a scholarly or higher education audience. In Calisphere, users reach the digital representations by searching or browsing through themed collections. This site is aimed at a very fully visualized researcher audience, primarily K-12 teachers and students. This pair of sites provided the most direct comparison on varying contextualizations of the same records.

Institutional repositories containing some archival materials:

- Scholars Archive at Oregon State University  
(<http://ir.library.oregonstate.edu/dspace/index.jsp>)
- Scholars Bank at the University of Oregon (<https://scholarsbank.uoregon.edu/dspace/>)

Institutional repositories (IRs) at colleges and universities offer long-term access to the intellectual output of the community, from student and faculty works to administrative records. They offer a particular example of long-term records and publications access for any type of organization. This is also another forum in which archival materials may be mixed with other, non-archival materials. These questions offered experience in two different IRs and allowed comparison of content and presentation as well as some assessment of the relative importance of IRs and similar types of archival programs.

Whole collections of archival materials linked to finding aids:

- Archives of American Art (<http://www.aaa.si.edu/collectiononline/>)
- Polar Bear Expedition Digital Collections site (<http://polarbears.si.umich.edu/>)

There is currently a strong movement in the archival profession to digitize entire collections rather than selected items. After ten or more years of archivists heavily selecting and/or contextualizing digital collections, this is a major change. Do researchers prefer to see whole collections, or would they prefer selected materials? A second part of this site pairing concerns the presentation of materials linked to the finding aid, essentially replicating the research room experience in the online environment.

Selected archival materials:

- Washington Women's History Consortium (<http://washingtonwomenshistory.org/>)
- Oregon State University's Best of the Archives  
(<http://digitalcollections.library.oregonstate.edu/cdm4/client/archives/index.html>)

Most commonly, archives create digital collections around topics or themes, selecting materials from larger collections according to some stated or, usually, unstated criteria. Multi-institutional sites can focus on a specific theme and present materials from across a state or region that relate to that theme. Single institutions often choose to showcase their "best" or most frequently used materials, often determined by researcher use or requests. The common assumption is that presenting a representative selection of materials will suggest to researchers what is available and pique their interest in pursuing other materials at the same institution. This set of questions assessed whether that assumption was true and tested satisfaction with key elements of these sites.

Archival materials: two different presentations of the same visual materials collection:

- On NWDA site: <http://nwda-db.wsulibs.wsu.edu/findaid/ark:/80444/xv10269>
- On the University of Washington's CONTENTdm site:  
<http://content.lib.washington.edu/19thcenturyactorsweb/index.html>

In this case, the same collection (the University of Washington's 19<sup>th</sup> Century Actors Cartes des Visite Collection) was presented with two very different major access points: an EAD finding aid in NWDA with links to the digital content from item-level lists at the component level, and a CONTENTdm collection with item-level records.

### ***Experimental Assignment***

We guided subjects in viewing three pairs of sites, allowing approximately twenty minutes to explore each site. Subjects were assigned sites based on the Latin Squares method. This technique randomizes the experimental sites shown to subjects as well as the order in which the sites are presented. This reduces potential learning effects or any other issues that might arise from viewing sites in a fixed order.

Table 2: Subject Assignment to Sites and Order of Presentation

Subject #	1: OAC/ Calisphere	2: WorldCat /Flickr	3: WWHC and OSU	4: 19th Century Actors	5: Institutional Repositories	6: Whole Collections
1	1	2	3			
2		1	2			3
3		3	1			2
4	3			1	2	
5	3			2	1	
6						1
7		2		1	2	
8				1		2
9	2				1	
10		1			2	3
11			2			1
12			1	2		
17	3	2			1	
18				2	3	1
20	1			4	3	
22		1				2
25	1			2		
27		3	1			2
29				3	2	1
Total Number of Subjects Viewing Site	7	8	6	9	9	10

As a result of this method we achieved randomization as well as fairly even coverage of all the sites. The unevenness occurred when subjects missed the phone interview and could not be rescheduled. Some interviews did not examine all three pairs of sites because of time constraints and very talkative subjects.

### ***Structured Interviews and Screen Sharing***

We conducted the 60-minute interviews in late summer and early fall 2008. The design of the study included hour-long interviews with the subject on the telephone while simultaneously using WebEx, an online conferencing application, for screen sharing. This enabled us to move through the protocol and project the sites, searches, images, etc. to the subject while ensuring that both were looking at exactly the same screen while talking on the telephone. The subject was then asked a series of questions based on the progression through the sites. We followed a scripted set of open-ended, directed, and ranking questions and a preset tour through a series of websites and activities on the sites. For our particular research questions this methodology worked well. There were some issues, including a time lag between the interviewer's search or scrolling action and the projection on the subject's screen, but these were resolved during pilot testing and through better timing of the script and checking with the subject frequently during the experiment to ensure his or her screen was in the right place.

### ***Analysis of Interviews***

We recorded the telephone portions of the interviews and these were transcribed. The transcriptions were then put into a text analysis system (AtlasTI) and coded for analysis. Initially, two of us separately coded the same interviews to ensure inter-indexer consistency. After several rounds of coding and discussion, we achieved a high degree of reliability. We developed several dozen codes, many coming from our research questions and others arising out

of the interview transcriptions. The codes fall into nine major categories. Some of the coding allowed us to sort other codes into categories based on the type of site or level of experience of the researcher. Other codes describe the researchers' reactions to the sites. These are outlined in Table 3 along with sample codes that fall into each category.

**Table 3: Coding Categories and Codes**

<b>Category</b>	<b>Sample Codes</b>
Experience with Site	Extensive Moderate Little None
Product	Grant proposal National Register application Book
Description	Amount of metadata Desired metadata elements (geographic, subject, copyright)
Search Entry Method	Keyword Browse Themed collections
Search Results	Resolution Keywords highlighted
Functions	Download images Contact an archivist Tag materials
Site Types	Format integrated

	Archives/special collections specific Selected from collections Whole collections
Credibility	Authority of information Reliability Moderation
Work Process	Strategies for research Methods of searching for materials

Once coding was complete, we proceeded to the analysis and identification of patterns relating to our research questions. These are discussed in the following section.

**Findings**

We grouped our findings into the three categories of the research questions: 1) Current use of Sites, 2) Metadata, and 3) Credibility.

*Current use of Sites*

Subjects' current use of websites that offer access to archival materials was tightly bound up in their personal research processes and interests. There was a high reliance on web-based archival information and a desire for more information and digital images on the web. This section will begin by contextualizing work processes and products and then discuss specific use of common features in sites (e.g. search, browse, etc.).

Comments about work process included strategies for approaching the archival websites and beginning a project:

I usually go in with somewhat general inquiries and then just kind of doing a survey of what's in there to help me decide whether or not I want to delve further but it's all predicated on the image not so much the text but more the image (Subject #5, line 121:125).

Every project that I start I sit down at my computer and see what is online first and that should tell how I work because if I can do more work at home on my computer, more research, then that saves me a lot of time and then I start and get into the car and I go to the archives. I go to the [name of institution]. I go to the library. I go to the museum and that sort of thing. My starting point is always at home online so the more information available in that format the easier it is for me. (Subject #1, line 546:551)

Participants also commented on the iterative nature of research. One genealogist (Subject #29) noted,

One thing leads to another ... the more research you do then you realize oh my gosh I've got to find out about that particular event and learn more about ... some piece of legislative history or some ... decision about ... where the railroad was going to go then you, it always takes you somewhere else where you didn't think you were going to go (lines 89:91).

Participants saw the virtual replication of reading room processes, such as the hierarchical access through a finding aid structure and a screen layout that was similar to flipping pages, as aligning with their work processes. As noted in the previous quotation from Subject #29, research is not linear. Returning to previous pages viewed and skipping around were cited as functionalities to support work processes.

One researcher went so far as to imagine his work process in a future with more archival information online:

It would be nice to be able to sit at my desk and view some of these materials and then work on getting copies, otherwise it's sort of a shot in the dark and can become quite expensive. But this way you could look and see and then say yes, this is something that would be significant, and I would like to get a copy of it if I could (Subject #10, lines 265:269).

The strongest and most numerous types of comments called for sites to be fairly comprehensive, offering access in some form or another to as many materials as possible regardless of format. Subjects desired context for materials, whether in the collection, geographically, or with links to related materials. They mostly described a need to gather comprehensive material in order to understand what was available and had a high expectation that digital sites could help them do that. Participants described a research process that usually moved from very broad to specific. Other comments included confusion over the purpose of institutional repositories, a general comfort with the finding aid as an entry point, and a desire to have the online environment

closely parallel the in-person research experience in terms of information available, visual cues, and available services.

Along with work process, subjects were very goal oriented in terms of their desired work products. Many comments focused on academic teaching and study. These included both use by the instructor in teaching or student use in the K-12 environment. In both of these cases, participants were most interested in illustrative images. Many subjects indicated that their work required access to images of facilities and buildings. For others, this was related to historic preservation, design, and design and planning work. These interviewees were interested in then-and-now images of buildings and landscapes, the ability to research a neighborhood or geographic area, building details, context, and access to blueprints and maps. Subject #4 observed that "Photographs would be good...I'm always comparing what's there now, what was there before?" (lines 169:173). Avocational users were interested in many of these same things to do detailed reconstructions of built landscapes.

Functionality is a key aspect in all archival websites. Perhaps no functionality is more central than navigation, which includes search, browse, and any thematic approaches to materials. Not surprisingly, subjects wanted to be able to easily and clearly find what they were looking for and preferred keyword searches as their entry point. At least two subjects observed that browsing lists are very subjective and dependent on someone else's preferences or perspectives. They also noted that browsing lists were inconsistent or didn't meet their interests.

The only thing I can think of is what I said before about whoever does the indexing on that A to Z should be as creative as possible and even if it is duplicative. You know put it under accidents and disasters. That sort of thing so that .... 'Cause we all think differently and it's like a guessing game so if you want to make it as useful as possible then let a couple of people index it. People come at it from different ways and come up with all the possible terms (Subject #9, line 445).

Themed collections were often frustrating to subjects and were perceived as most useful for teachers or students, "I think it's probably good for more amateur users...there are a lot of people that don't do well at formulating keywords and getting what they want. They don't quite have the knack, the language, so providing a browsing list I think is very helpful for them (Subject #1, line 611).

In terms of search results, participants had a very strong desire for an obvious connection between the search terms they used and the results the site presented; fourteen subjects offered comments to that effect. Particular features mentioned were a keyword-in-context display, highlighting search terms in results, and presenting connections at the top of the screen or in a new screen. Subjects were displeased with most examples of keyword displays because they could not discern whether the text was taken from the object itself or its accompanying metadata:

... we know that we're looking at five riverbed miners but if I didn't know ... if I hadn't seen this picture already, I wouldn't know if that was the actual title of the photo or if was just a description of it or whether or not that's relevant. I wouldn't know anything

about this. (Subject #20, line 419)

Others had trouble with presentations that they found too dense and difficult to read:

I liked the fact that “miner” is highlighted in the text but I think the search terms and content is a bit dense in terms of text and I can understand what they’re trying to do with that but maybe there’d be another way to do it because it’s really not . . . with all the ellipses in there, it’s not terribly illuminating. (Subject #5, line 19)

Closely connected with this, subjects asked for a specific expression of how results were ranked, including options for changing that ranking easily. “I’m curious how they get ranked, how they are prioritized.” (Subject #22, line 275)

Nine subjects also wanted results to be presented both visually and with some accompanying text to be able to view results both ways. Participants self-identified as visually- or textually-oriented, with approximately half in each group. For textual materials, visual presentation was useful to discern format (the search on the Polar Bears site was specifically set up with a form-genre component that could be resolved visually), but subjects also wanted textual cues. The types of text information that they found most desirable included the item title (if it had “a bit of meat”) and a description or scope and content note. Subjects were frustrated by repetitive data presented on some sites, and some questioned whether the name of the holding institution was all that important if they expected to complete their work online rather than as preparation for an institution visit. Regarding visual materials, the subjects who indicated that they were visually

oriented, or those who were most likely to be searching for visual materials, had specific visual qualities in mind, including composition, orientation, or format. Thus they found it highly desirable that sites enable visual scanning.<sup>7</sup>:

Well I like the fact that it is kind of like Ebay in a way that not only does the two tell the titles and what is there but also a little visual icon, I guess. Those photos, what it looks like? So I think that really helps if you are trying to zero in on something in particular. And if you see something in the photo that looks like something you might be able to use. For me that would be helpful going right to that particular document instead of looking at things you don't really need to see. (Subject #4, line 205)

However, they also said that visual results without any text (as they are presented in one area of the OAC) offered insufficient information for selection. Subjects also had very specific desires or criteria for evaluating the size of images and text and were particularly sensitive to sites that presented small images in a sea of white space they felt was wasted. Some were aware of or receptive to changing the screen's appearance by making browser adjustments, but not always. As noted elsewhere, the presentation of the covers of textual objects or objects with elements like tintype case hinges as the image thumbnail was universally confusing: "Why is there a red box for the first one and nothing on the right hand side of the screen for the other two?" (Subject #1, line 203).

---

<sup>7</sup> This is consistent with the findings in Kathleen Fear, "User Understanding of Metadata in Digital Image Collections Or, What Exactly Do You Mean By 'Coverage'?" *American Archivist* 73/1 (Spring/Summer 2010): 26-60.

Subjects were very clear that they wished to review only results sets of a small to very moderate size (up to about 20 items) with textual and visual presentations. Seven subjects said that they desired useful limiters or sort functions for search results to reduce the need to scan very large results sets. Participants also stated that they got ideas for appropriate limiters from the ways in which their search terms were presented in or connected to the results set. They indicated that while two of the searches performed on the sites were satisfactory to review without limiters, it was only because there were less than ten items in the results sets. If the set had been larger and without useful limiters, they would have given up and abandoned their searches. The types of limiters they desired included keyword and form and genre. Two subjects also mentioned that they preferred interfaces like Microsoft's photo editor that allowed them to simultaneously see many small images and one large image on the screen at the same time.

We were also very interested in exploring which functionalities best supported researchers work processes and would most facilitate their work. We either raised questions about, or subjects mentioned sixteen different functionalities (aside from search and browse). The discussion of these items will be divided into a discussion of traditional functionalities and Web 2.0 features. A full list is presented in Table 4.

**Table 4: Ranking of Functions by Subjects**

Function	# of Subjects	Positive/High	Neutral/Moderate	Negative/Low
Get High-Quality Images	14	14	0	0
Use Others' Tags	12	11	0	1
Read Comments by Others	16	10	2	4
Contact the Archivist	12	9	2	1
Leave Comments for Others	14	7	4	3
Linking	7	7	0	0
Zoom for Images	11	7	1	3
Download/Save Files or Images to Desktop/Favorites	7	6	0	1
Sort	6	6	0	0
Print Screen	5	5	0	0
Search OCR Text	6	5	1	0
Tagging	10	4	0	6
Look at Oral History Transcript	4	4	0	0
Listen to Oral History	3	3	0	0
Contact Other Researchers	9	0	4	5
Social Software	4	0	0	4

### Traditional Functionalities

Subjects wanted access to the archivist and relied on the archivist for information. Twelve shared comments on the ability to contact the archivist. In general, they were quite a bit more interested in contacting archivists than they were in contacting other researchers: “If I have a specific question about something, I would be more inclined, like about the processing of it or whatever, to contact the archivist” (Subject #5). However, several subjects indicated that they would contact an archivist not so much for in-depth knowledge of materials, but to obtain copies: “The other thing . . . on any archival site is how accessible is the archivist and the research staff, I guess, if you’re researching from afar and you need copies of things. . .” (Subject #17, line 71)

The ability to get high-quality reproductions of images was the most frequently mentioned functionality. Fourteen subjects commented on the ability to get high-quality images for publication or study, with every one of them expressing a high interest in this function.

Additionally, participants very much desired a streamlined process for doing so:

Right now we have to copy the form down and you can put your requests on there and figure out from their price how much you owe them and you can call and give them a credit card. You know Amazon or Ebay or something where you could actually go in and select and I realize this wouldn't work for everything but even if you could make your selections and provide your purchasing information and then work out the details.

That would be incredibly useful for us. (Subject #12, line 245)

While most focused on a more traditional process of placing an order for prints or photocopies, there was interest in the ability to download high-quality images directly for further study rather than for publication: "I want the maximum resolution on a print and yes, I'll pay the 30 bucks for it" (Subject #11, line 443).

Subjects were keen for archives to take full advantage of straight HTML. Linking between collections and being able to sort search results were on the top of this list. The most highly desired links were ones for context that provided connections with materials in other formats, including museum objects, with the same creator or subjects: "I would like to know is there anything associated with it, like are there any oral histories; are there stories; are there documents

associated with it that might help complete a larger story” (Subject #12, line 71). Subject terms were the next most desired types of links. Interviewees saw sorting as a means of making reviewing results more efficient: “I don’t know if this has a sort option or not but you can sort, for example, show me all the images or show me all of the publications without the images or show me all of the primary documents” (Subject #1, line 279).

Participants also expected other simple functions that they find on generic websites: printing the screen, downloading and saving files or images to their desktop or favorites file, and zooming for images. Subjects commented that they cannot read archival materials directly off the screen and/or prefer to amass personal collections for further study. They also desired the ability to print entire documents rather than a page at a time. When viewing images in our study, subjects commonly asked, "Is there an ability to save this image and to collect say similar images in a file, separate file, on somebody’s computer?" (Subject #1, line 75). Zooming was discussed in reference to both visual and textual images, particularly for close image study for historic preservation: “Almost unnecessary but I really like it. The detail on this is amazing. It seems like they put most of their work, their focus on getting really high resolution pictures for use” (Subject #25, line 126).

### Web 2.0 Features

Subjects were wary of the Web 2.0 features offered on these sites. Despite the high ranking of three features (using others' tags, reading others' comments, leaving comments for others), there were also a number of negative or at least wary comments. Since we were specifically

interested in this, we also often had to ask people about their opinions of the use of Web 2.0 on archives websites during the experiments. Few subjects raised this issue on their own.

Participants were more interested in taking advantage of information left by others than in contributing their own information to archival websites. Most were individual researchers and few were genealogists, so they did not see themselves as part of a larger community. This may be the reason for this tendency. Twelve subjects discussed using others' tags, with eleven indicating interest in this feature but not necessarily for themselves: "I think tagging would be useful because some of the stuff goes way back and you don't know the wording, what they are using. . . that broadens the searchability" (Subject #18, line 401). But participants were just as likely to question the validity of tags:

I think those are just goofy. I don't think somebody doing research is going to . . . if you're trying to do serious research, you're not going to click on that. If you're looking for something on the Gravelly Range or Madison County, well, even Madison County, it's going to take you to how many Madison Counties so if you're looking for a particular Madison County that's not going to take you necessarily to the only one that you're looking for. I don't think they're that reliable. I think they can probably help you find some things but your example of sheltie, collie shows you the downside of social tagging and what's going to come up if you click on transparency? You don't know. (Subject #3, line 373)

Sixteen subjects discussed reading others' comments. Of those, ten were positive, two neutral, and four negative. The range of comments tells the story: "I'd love to be able to search

comments . . . Any data is useful data” (Subject #6, line 275) and “I will be honest. Most websites I go to, I sometimes read the comments, usually I don’t. That is not as important to me as the actual detailed information that was above that” (Subject #10, line 161). Subjects also noted that sites tied to more focused user communities tended to generate considerably more useful comments than general sites like Flickr or WorldCat.

In spite of the interest in partaking of others’ tags or comments, there was less interest in tagging or commenting. Of the fourteen subjects discussing this, the communication method shaped their responses; many felt that they would be unlikely to post a comment, then return to a site to look for responses: “I wouldn’t post because I wouldn’t want to follow up and have to go back and get the answer” (Subject #2, line 111). Time management was also an issue: “I just usually go in with a specific task and try to get it done and if I have questions I’ll contact somebody later” (Subject #5, line 335). However, subjects sometimes framed their comments in terms of the perceived needs of others rather than their own needs. Of those who desired this, several mentioned the importance of being able to search comments for added metadata and the importance of sharing knowledge in this way, but generally preferred to seek expertise in other arenas. Tagging generated a similar response, and also “There’s no description so if tagging can help with that, that would be nice” (Subject #7, line 159).

There was also little desire to enter into relationships with others using a website. Subjects stated that they preferred to use other methods like posting to specific listservs where people with relevant expertise would be likely to answer and point them toward appropriate resources. Most of them said that they wanted to know who they were contacting and/or had comments about

reliability. Another expressed her preferences in terms of time: “I don’t join chat rooms. I don’t know how people have time for that. . . I don’t. . . I kind of reached the point where I have to really focus on what I’m doing and whether time is invested in things like what you’re suggesting and contacting others is really time well invested” (Subject #17, line 136). The researchers interviewed did not seem to identify with any of the communities on the sites: “If these were users that I knew then it would matter” (Subject #27, line 432).

In general, the subjects were most interested in functions that supported traditional research activities: contacting archivists for more information on collections or publication permissions, getting copies, and links or other connections to related materials. They wanted sites to make these tasks faster and more convenient with features like online ordering and paying that they have come to expect from their experiences with commercial sites. Participants were less enthusiastic about other, mostly Web 2.0, functions. While some saw commenting and tagging as useful as ways of expanding available metadata, few indicated that they saw these types of sites as places to meet other researchers and exchange information; most had other established modes for those functions. As in other areas, subjects were quite sensitive to questions of expertise and credibility and perceived that unless they had a way of knowing the knowledge level of others, they would not trust their comments or tags.

Last, subjects’ expectations for what sites should offer and how they should function were clearly shaped by their experiences with commercial sites, with four subjects discussing this. Not surprisingly, they specifically mentioned Amazon and Google. They wanted results presented with a combination of visual and text elements and the ability to choose how many results they saw on the screen at a time.

## *Metadata*

Description and metadata issues are key issues for online archives. This was a key area of interest, and one in which we received some of the most substantive comments. Subject #17 articulated the importance well: “The descriptions in archives sometimes are written with the assumption that the researcher is familiar with archives, you know, the terminology and the length of the description or whatever and researchers are in various stages of their knowledge of that” (line 39). In addition to the value of different types of metadata, one of our research aims was to better understand the level of detail subjects needed in descriptive metadata. Thus, our site selection was designed to draw out these types of comments. This section begins with a general discussion of ‘level of metadata’ and then moves into a discussion of what participants said about particular metadata elements.

Sixteen out of the 19 subjects made a total of 88 comments about the level of metadata. Not surprisingly, they wanted as much detail as possible. The pragmatic issue then becomes the difference between what subjects *want* and what subjects *need* to make selection decisions about results screens. Throughout the experiment, subjects were asked to comment on the descriptive metadata in different ways. These ranged from asking for general commentary on the metadata to inquiring whether they would be able to make a selection decision based on the information presented on the screen.

When asked, subjects almost always wanted more information about collections and items. The need for more information seems to be more acute at the item level with digitized archival

materials, particularly photographs. Several subjects also mentioned a desire to have multiple levels of description available, a brief view linked to a fuller description. In the words of one participant: “It has the same idea with the thumbnails and the brief description and the ability to go to the details from there. I think by providing that to people to scan first before selecting is a good way to present it” (Subject #1, line 127).

Subjects also expressed a desire for item level metadata on visual images. This latter finding goes against recent initiatives to limit item-level metadata during digitization and “more product, less process” materials preparations. Sometimes the metadata they wanted was extremely detailed. Subject #7 noted that “it is often addresses or street intersections or neighborhoods” (line 223) that were the important geographical features. Additional metadata is an issue for selection as well as interpretation. Having this information can mean the difference between using a photograph and doing an additional search:

Subject #12: I think had I gone the other way and then found that photo and nothing came up with it, I would have assumed there wasn't anything else to be gained. I wouldn't have necessarily gone back and tried another avenue to find information about it. Unless it was a really hot photo and I might have called them up. Then I would have pursued it more...

Interviewer: So having information like this might mean the difference to you between not using a photograph and seeking permission to use a photograph about something.

#12: Right, because depending on how anxious you were to get that photograph I might have moved on to one where I could see there was information instead of working harder at it to try and figure it out. (Line 229:233)

This exchange also shows that archives are in competition with other sites and researchers will only exert a certain amount of effort before they decide to move on.

As previously noted, when viewing several of the test collections (Online Archive of California, Calisphere, Polar Bear Expedition Digital Collections, University of Washington 19th Century Actors, Northwest Digital Archives, WorldCat, and the Best of the Oregon State University Archives) subjects were asked if they could make a decision based on the metadata provided in a search results set. The test was somewhat problematic because many of the interviewees had trouble imagining themselves pursuing research questions similar to the ones we posed. Still, the results were chilling; there was no site where a majority of participants said they had sufficient information to make a selection. The major reasons cited by the subjects were an absence of the search terms or a good description in the results and the size of the thumbnail image. This finding applies primarily to the photographic collections in the tests and points to the importance of good labeling and some apparent indication of why the results set was retrieved. For example, in the University of Washington and the Northwest Digital Archives 19<sup>th</sup> Century Actors sites, a search for ‘minstrel’ retrieved a number of images but it was unclear from the initial metadata whether the individual portraits were of minstrels or whether ‘minstrel’ appeared somewhere else in the text and had nothing to do with the subject. The gap between existing metadata and what people need to make decisions about identifying, selecting,

interpreting, and using online archival materials needs to be explored in greater depth. The crucial question becomes not what users want, but what they need.

**Table 5: Ability to Make a Selection Based on Metadata**

Site	Yes	No	Uncertain
Best of the OSU Archives	1	1	1
Calisphere	2	2	1
Northwest Digital Archives	1	1	0
Online Archive of California	1	1	2
Polar Bear Expedition Digital Collections	1	3	1
University of Washington 19 <sup>th</sup> Century Actors	2	5	1
WorldCat	1	1	2

Metadata Elements

Subjects discussed 20 separate types of metadata. The most frequently cited types of metadata were geography, scope/content information (summaries above the item level), subject, dates, copyright, and document type/genre. Figure 1 provides a graphic illustration of the number of subjects citing each of these metadata elements.

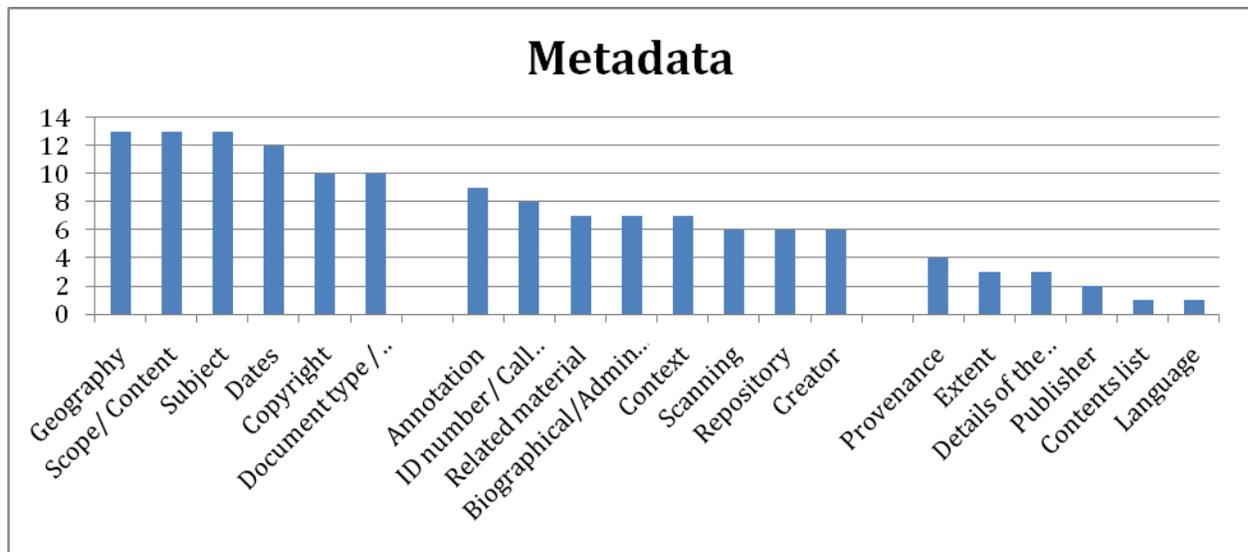


Figure 1: Metadata Elements Cited by Subjects

The frequency of citation only tells part of the story. How subjects talked about these elements and their importance, their impressions of the information and their terminology, are all important in understanding how to make archival resources more readily discoverable. In the words of one of the interviewees, “Archives contain an overwhelming amount of information and usually I’m looking for something particular” (Subject #17, line 59). This list of metadata elements contains some expected responses, such as the desire for additional geographic and subject access. When shown more sophisticated geographic information system (GIS) features, however, only two (#22 and #27) of the seven participants who were exposed to the Library of Congress’ Flickr site commented on the map visualization as an access point. Both were impressed and thought this held possibilities for other sites. One envisioned using a GIS application to search for collections on his own:

You know what would be really cool is if there were - this is always my dream - if there were a link, like a GIS link or even in a very general sort of way to where the Motherlode is and you could save that or something like that and then save it with your favorites and then at the end of the session you could see where all of these different favorites have occurred or something like that. With all the technology now, that kind of thing should be really easy to do. It doesn’t have to be, like I said, terribly specific (Subject #5, line 64)

Subject access was very popular. The sites that provided easy and visible subject linkages (OSU, Washington Women’s History, even Flickr) were very popular in this regard even though participants were somewhat ambivalent about the tags on Flickr.

Copyright emerged as one of the most important metadata elements, with ten subjects discussing it. The popularity of copyright goes along with the desire of most of the subjects to be able to download, edit, and manipulate text and images. Yet, this desire is tempered by copyright: over half of the subjects were attuned to intellectual property issues and wanted clear information on websites concerning re-purposing of materials. While none of the participants said that he or she would not use a photograph if they could not get permission, copyright and discussion of potential use were often mentioned hand-in-hand.

Key archival elements: creator, repository, and provenance were some of the less desired metadata. Six subjects mentioned the desire to have information on the creator. Interestingly, all of these subjects discussed the need for creator information **only** during photographic searches (Washington Women's History (3), Best of OSU (1), OAC (1), NWDA (1), Flickr (2)) and on sites where no creator information was listed. When it was there, it apparently became part of the woodwork and not worth mentioning. Only Subject #1 was confused by the meaning of author or creator in the representation of photographic images: "Tell me what the author means, the author field means in this particular image (line 59 in relation to Washington Women's History) and I don't know what creator means. I think that's confusing to me. Does that mean photographer? (line 131 in relation to Best of OSU)".

Somewhat surprisingly, the need for more transparent repository information occurred on the union sites (OAC, OCLC, Washington Women's History Consortium) as well as single repository sites (UW – 19<sup>th</sup> Century Actors; LOC Flickr). Although the number of times

‘repository’ was identified as important metadata is low, conclusions cannot be firm, but because this element was considered important on both sites where the repository was obvious and where it was not, it does appear that this is an expected element.

In addition to being able to identify the repository, four subjects wanted to know the origin or custodial history of the archival materials. This was linked to both interpretation, “I’m going to be curious whether it’s a Kodak or a professional photographer so I can figure out where it came from” (Subject #11, line 435) and additional resource discovery, “One thing that I do like is under the notation...transferred from the US army, US Office of War Information. That gives me an idea of sort of its provenance, and I can think OK, maybe I can look under other stuff from the office of war information, and see what else is out there” (Subject #10, line 149). Two of the three subjects who mentioned the details of the physical object viewed the Online Archive of California and reacted to the “Image package note” metadata element. Two subjects mentioned publisher and one each mentioned a contents list and language of the materials. It was unclear exactly what the two subjects meant when they identified publisher as a desired metadata element – this could be the original publisher of the photograph or the repository as the online publisher.

### ***Credibility***

Credibility and authority are increasingly important as more archival materials are digitized and put online. Thirteen subjects commented on credibility issues. These comments were unevenly divided between wariness about the reliability of information from social computing features (tagging, annotation, commenting) and the reliability of the archivists’ official descriptions.

Interestingly, social computing features were not as popular as anticipated. Most of the subjects discussed credibility in relation to social computing. One subject provided a very pragmatic reaction to tags: “Sure but you could always, you know, figure out if it is valuable or not .... I think it is easy enough to check it out. I wouldn’t be worried about it...I might not get on a plane and go somewhere based on them” (Subject #29, line 442). As noted elsewhere, subjects did not rule out social tagging and commenting. Also, at least one participant felt that the onus was on the other site visitors and not the archivist to vet crowd-sourced information:

I think as a participant in both academic and in just internet usage that you sort of need to develop a sense of can I trust this source or not. That only comes through experience. I don’t think you can blanketly say that internet commenting is either all right or all wrong and I think that is up to the users to determine what they want to use and how they’re going to use it. I think just the availability is worth the possible inaccuracies. (Subject #20, line 219)

While they mostly wanted to contact archivists for credible information, subjects also expressed some skepticism about archivists as credible sources. Two commented on official archival descriptions that they noted were sometimes inaccurate. Also, in spite of the desire for both item level annotations and broader scope and content information, subjects were wary of information, particularly annotations, titles, and captions that had no apparent source. This was especially important for image captions when subjects could not tell if they were supplied by the archivist or transcribed from the verso of the photograph. As noted by Subject #11: “The quote I assume

is, it doesn't say, it's apparently from something on it, it doesn't say that, it's an assumption. If that's what's written on the photo I'll go with that as opposed to what someone thinks of it" (line 335)

## **Discussion**

No archival system or website can fully support researchers' iterative, non-linear search processes. But this study points to some steps that can be taken in this direction. These include better search functions, a focus on supporting traditional functionalities such as up-front contact information for the archivist, as well as increasing support for remote users in terms of downloading and using images (copyright) over Web 2.0 features, and enriched metadata that provides both content and context.

Subjects preferred keyword search over browsing categories or pre-established themed collections. They saw keyword search as enabling them to begin comprehensively and then move on to either sorting or narrowing large results sets. Interviewees were dissatisfied with browsing lists and themed collections, preferring more control over their search environment.

Overall, subjects expressed interest in traditional reference functions more than they wanted web 2.0 functions. In general, they preferred an online environment that closely paralleled the on-site in-person research experience in terms of information available, visual cues, and available services. This may be an artifact of the individual research practices that most followed. Had there been more genealogists in the sample, the results might have been different. Functions of greatest interest included easy online access to the archivist for more information or permissions,

a streamlined process for procuring copies for study or publication, and the ability to easily print or download items for study.

Our participants found that none of the sites used in the study gave them sufficient descriptive information to select items. This may have been a drawback in our methodology. It would be important to replicate this part of the study with a slightly amended methodology and have subjects pursue their own questions so that they were more knowledgeable of the purpose of the selection and could better judge the value of the metadata.

In discussions of desired metadata elements, subjects cited geography, scope/content information (summaries above the item level), subject, dates, copyright, and document type/genre more frequently. Interestingly, participants also wanted contextual material. When given a chance to compare the presentation of the same collection in NWDA or a digital asset management system, many subjects preferred the finding aid view because of the improved collection-level and contextual information available there. Subjects' desire for more metadata is clearly at odds with fiscal reality in most archives.

The issues of Web 2.0 and metadata are tightly bound up with issues of credibility. Participants' entire discussion of the value of Web 2.0 features revolved around the credibility of the author of the tags or comments and the reliability of the information. While subjects did have interest in reading others' comments, they were very wary of relying on these without verification. On the other hand, they also questioned the official archival descriptions in two distinct ways. In search results, subjects wanted to know if the term retrieved was original to the item (e.g., a photograph

caption) or a term supplied by archivists. In image collections, the interviewees were concerned about the origin of captions, and whether archivists had transcribed a note on the back or supplied the information themselves.

There are clearly tensions between both the ability of current systems to deliver content as the subjects would like and in the archival management systems to process collections to the degree desired. However, in this small sample, subjects did not see crowd-sourcing as an attractive way to bridge the gap.

Finally, it is a given that researchers want more materials available online. Yet, few of the subjects had used any of the sites in this experiment. This raises the issue that researchers are not aware of many of the sites that do exist, and that there is no one place to go to search all of the archival materials online, nor even any union list of sites. Thus, researchers are not taking full advantage of the existing online archival materials. How closely this is related to their research habits or to the archival community's slowness to embrace user studies merits examination.

## **Conclusions**

This study was propelled by NWDA's and the Orbis Cascade Alliance's desire to develop a sustainable digital services program for presentation of locally held unique materials, primarily those in archives and special collections, with their memberships. In other portions of the needs assessment and planning for this program, NWDA and Alliance members have expressed both a desire to know their audiences and a frustration at the lack of knowledge of those researchers'

needs. Many existing digital programs or projects were created and shaped with limited knowledge of the needs they would meet, and efforts to gather and use qualitative or quantitative data on their use have been inconsistent. Although the results of this study are preliminary due to the limited number of subjects involved, the program has found them sufficient to be useful. Additionally, they are quite congruent with the findings of the (currently unpublished) Mellon Foundation-funded user study of the Southern Historical Collections.<sup>8</sup> Naturally, as resources contract rather than expand, organizations must raise questions about the degree to which they can continue or increase support for digital programs. For any institution creating or reforming a program to present digital content online, creating a program that meets the needs of core audiences is essential for its long-term success.

## References

Choi, Y., & Rasmussen, E. M. (2003). "Searching for images: The analysis of users' queries for image retrieval in American history." *Journal of the American Society for Information Science and Technology*, 54(6), 498-511.

Coats, L. R. "Users of EAD Finding Aids: Who Are They and Are They Satisfied?" *Journal of Archival Organization* 2, no. 3 (2004): 25-39.

Czeck, R. L. H. "Archival MARC Records & F.A.'s in the Context of End-User Subject Access to Archival Collections." *American Archivist* 61 (Fall 1998): 426-439.

---

<sup>8</sup> For information on the outcomes of this initiative, see <http://www.lib.unc.edu/mss/archivalmassdigitization/index.html?section=introduction>.

Fear, Kathleen. "User Understanding of Metadata in Digital Image Collections Or, What Exactly Do You Mean By 'Coverage'?" *American Archivist* 73/1 (Spring/Summer 2010): 26-60.

.

Foundations Project. (2000). "Foundations Project usability testing: Dublin Core metadata and controlled vocabulary study." Retrieved September 13, 2008, from <http://www.bridges.state.mn.us/user2study.pdf>.

Hamburger, Susan. "How Researchers Search for Manuscript and Archival Collections." *Journal of Archival Organization* 2, no. 1/2 (2004): 79-102.

Howard, Dawn E. The Finding Aid Container List Optimization Survey: Recommendations for Web Usability. A Master's Paper for the M.S. in L.S degree. November, 2006. University of North Carolina.

Hutchinson, T. B. (1997) "Strategies for Searching Online Finding Aids: A Retrieval Experiment." *Archivaria* 44 (Fall 1997): 72-94.

Jorgensen, C., & Jorgensen, P. (2005). Image querying by image professionals. *Journal of the American Society for Information Science and Technology*, 56(12), 1346-1359. doi: 10.1002/asi.20229.

Kim, J. (2004). EAD Encoding and Display: A Content Analysis. *Journal of Archival Organization*, Vol. 2(3), 41-55.

Library of Congress, American Memory User Evaluation, 1991-1993,

<http://memory.loc.gov/ammem/usereval.html>

Liddy, E.D., Allen, E.A. & Finneran, C.M. (2003). "MetaTest: Evaluation of Metadata from Generation to Use." *Proceedings of the 2003 Joint Conference on Digital Libraries (JCDL '03)*.

Retrieved September 13, 2008 from

<http://csdl2.computer.org/comp/proceedings/jcdl/2003/1939/00/19390398.pdf>.

Northwest Digital Archives, Executive Summary: Usability Testing Round 4

Compiled by Tiah Edmunson-Morton, UTWG chair, 2008-03-12.

[http://orbiscascade.org/index/cms-filesystem-action?file=nwda/reports/nwda\\_utwg\\_ut4\\_report\\_20080312.pdf](http://orbiscascade.org/index/cms-filesystem-action?file=nwda/reports/nwda_utwg_ut4_report_20080312.pdf)

Northwest Digital Archives, Usability Design Working Group, Executive Summary

Usability Test #5 – Search Functionality 2008 November 6, [http://orbiscascade.org/index/cms-filesystem-action?file=nwda/reports/udwg\\_ut5\\_executive\\_summary\\_20081106.pdf](http://orbiscascade.org/index/cms-filesystem-action?file=nwda/reports/udwg_ut5_executive_summary_20081106.pdf)

Northwest Digital Archives, Survey of Digitizing Initiatives Report. Compiled by Jodi Allison-

Bunnell, NWDA Program Manager, 2007 November 20. [http://orbiscascade.org/index/cms-filesystem-action?file=nwda/files/di\\_survey\\_report\\_1107.doc](http://orbiscascade.org/index/cms-filesystem-action?file=nwda/files/di_survey_report_1107.doc).

Northwest Digital Archives, Researcher Type Survey Report. Compiled by Jodi Allison-Bunnell, NWDA Program Manager, 2008 January 23. [http://orbiscascade.org/index/cms-file-system-action?file=nwda/files/researcher\\_survey\\_narrative\\_20080123.pdf](http://orbiscascade.org/index/cms-file-system-action?file=nwda/files/researcher_survey_narrative_20080123.pdf).

Northwest Digital Archives, Institutional Needs Survey Report. Compiled by Jodi Allison-Bunnell, NWDA Program Manager, 2008 October 25. [http://orbiscascade.org/index/cms-file-system-action?file=nwda/files/institutional\\_needs\\_survey\\_report\\_20081025.pdf](http://orbiscascade.org/index/cms-file-system-action?file=nwda/files/institutional_needs_survey_report_20081025.pdf).

Online Archive of California, OAC Usability Test Summary (October 2001).  
[http://www.cdlib.org/inside/assess/evaluation\\_activities/oac\\_usabilitytest2001\\_summary.pdf](http://www.cdlib.org/inside/assess/evaluation_activities/oac_usabilitytest2001_summary.pdf)

Online Archive of California, OAC Usability and Survey Results Summary, July 2002.  
[http://www.cdlib.org/inside/assess/evaluation\\_activities/OACUsabilitySummaryJuly2002.pdf](http://www.cdlib.org/inside/assess/evaluation_activities/OACUsabilitySummaryJuly2002.pdf)

Online Archive of California, OAC First Round Usability Test Findings: OAC Redesign Project, Prepared by: Jane Lee, 11 September 2008.

[http://www.cdlib.org/inside/assess/evaluation\\_activities/docs/2008/oac\\_usability\\_aug2008.pdf](http://www.cdlib.org/inside/assess/evaluation_activities/docs/2008/oac_usability_aug2008.pdf)

Online Archive of California, OAC Second Round Usability Test Findings

OAC Redesign, Prepared by: Jane Lee, 23 June 2009

Project [http://www.cdlib.org/inside/assess/evaluation\\_activities/docs/2009/oac\\_usability\\_april2009.pdf](http://www.cdlib.org/inside/assess/evaluation_activities/docs/2009/oac_usability_april2009.pdf)

Prom, C.J. (2004) *American Archivist* 67(2): 234-268.

Scheir, W. (2005) First Entry: Report on a Qualitative Exploratory Study of Novice User Experience with Online Finding Aids. *Journal of Archival Organization*, Vol. 3(4), 49-85.

Quam, E. (2001). Informing and evaluating a metadata initiative: Usability and metadata studies in Minnesota's Foundations Project." *Government Information Quarterly*, 18(3), 181-194.

Stebbins, Robert, *Serious Leisure: A Perspective for Our Time*. New Brunswick, NJ: AldineTransaction Publishers, 2006.

Yakel, E. (2004) EAD: Are Finding Aids Boundary Spanners or Barriers for Users?" *Journal of Archival Organization* 2(1/2), 63-77.