

Amanda L. Loeb. Shopping the Online Archives Megastore: A Content Analysis of Special Collection Libraries and Archives Websites Produced Through Large-Scale Digitization. A Master's Paper for the M.S. in L.S degree. December, 2013. 32 pages. Advisor: Jacqueline M. Dean

This study is a content analysis examining websites of special collection libraries and archives supporting large-scale digitization initiatives to determine the kinds of information and functionalities available to online users. Large-scale digitization is characterized by aggregate-level selection and metadata of materials to be digitized. The analysis was conducted on a group of ten institutions identified as participating in large-scale digitization. The results demonstrate success in highlighting digitized materials on the website, but show a lack of consistency in the use of item-level and aggregate-level metadata, suggesting inconsistency in the definition of large-scale digitization.

Headings:

Content Analysis

Special Collections

Digitization Projects

Archival Materials - Digitization

SHOPPING THE ONLINE ARCHIVES MEGASTORE: A CONTENT ANALYSIS OF
SPECIAL COLLECTION LIBRARIES AND ARCHIVES WEBSITES PRODUCED
THROUGH LARGE-SCALE DIGITIZATION INITIATIVES

by
Amanda L. Loeb

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

December 2013

Approved by

Jacqueline M. Dean

Table of Contents

Introduction.....	2
Literature Review.....	4
History of Mass Digitization in Libraries.....	4
Large-Scale Digitization in Archives and Special Collections.....	4
Archival Users.....	7
Methodology.....	9
Results and Discussion.....	13
Discovery and Access.....	13
Digitized Material.....	17
Metadata.....	20
Functionalities.....	22
Conclusions	26
Bibliography.....	29

Introduction

With the world now firmly entrenched in the digital age, archival and special collections materials are no longer confined to the library. Through digitization, archivists and librarians have overcome the geographical distance between researcher and repository, and united geographically disparate collections of rare and unique materials. Digitization is now an undisputed part of the archivist's work. With an increase in capabilities, however, comes an increase in expectations. Users have requested greater online access to archival and special collections materials. "Boutique" digitization projects that painstakingly curate, capture, and describe limited designer collections are no longer sufficient to meet the needs of users: they want to shop the online archives megastore. Archivists and librarians have responded by implementing large-scale digitization of materials, providing online access to large extents of material. The trade-off is the limited role of the archivist in selection of materials for digitization and presentation online, and the implementation of minimal descriptive metadata for digitized materials.

There are a variety of factors to consider when planning and executing a large-scale digitization initiative. How much will be digitized? What formats of materials can be digitized? What kind of equipment will be used? How will project managers ensure quality control? While there has been much discussion of the theory behind large-scale

digitization, justification for the approach, and workflows for implementation, there has been little discussion or analysis of how researchers use materials produced in large-scale digitization initiatives, or the features and capabilities typically included in the online interface. This study consists of a content analysis on websites of special collection libraries and archives supporting large-scale digitization initiatives, in order to better understand how users can interact with resources produced through the large-scale approach.

Literature Review

History of Mass Digitization in Libraries

While large-scale digitization is an emerging trend in archives and special collections, libraries and other institutions have more commonly used the process to digitize large holdings of books and bound materials. Many libraries have entered into corporate partnerships in order to accomplish digitization of books on an industrial scale (Weintraub, 2008; Ceynowa, 2009). In the case of the Bavarian State Library, librarians chose to partner with Google to outsource scanning of their copyright-free collections dating from the seventeenth to the nineteenth century. The selection of materials was based solely on copyright status and physical fitness for scanning in terms of conservation, size, and volume. Selection was in no way curated to focus on certain subject areas, authors, or other factors. Google retained rights to the digital copies, but also provided digital copies to the library for its own use. Librarians were free to provide users with access to the digital copies through the library catalog and website. Under several like projects, the Bavarian State Library brought over 1.2 million books in its holdings online.

Large-scale Digitization in Archives and Special Collections

In recent years, several special collections libraries and archives have undertaken mass or large-scale digitization projects in order to provide archival users with online

access to materials.¹ The goal of large-scale digitization is generally agreed to be providing access to larger quantities of resources at the collection level, rather than small amounts of digitized materials at the item level. Accordingly, most projects attempt to digitize whole collections, the bulk of a collection, or entire series (Chapman and Leonard, 2013; Sutton, 2012).

The current literature identifies selection of materials as a defining component of a large-scale digitization initiative. In large-scale digitization projects, the archivist performs minimal selection of material, instead opting to designate large quantities of materials from one or more collections for digitization without reviewing every single document. The decreased emphasis on selection in large-scale digitization allows archivists to focus on providing access to greater extents of material, whereas the time and resources required for “boutique” digitization often limits the scope of projects (Chapman and Leonard, 2013; Sutton, 2012). Archivists hold that this approach accomplishes the request of scholars and archival users to preserve the context of archival materials in the digital environment (Rieger, 2010). Large-scale digitization initiatives also support users’ expectations of being able to access large quantities of information via the web (Greene, 2010; Gueguen, 2010; Reiger, 2010). While there is some concern that decreased efforts in selection will increase the risk of publishing copyrighted materials, archivists have employed fair-use practices in an effort to provide as much access as possible to digitized material (Smith, 2012).

¹ Examples of large-scale digitization projects in special collection libraries and archives include the Digital Southern Historical Collection at the University of North Carolina at Chapel Hill (<http://dc.lib.unc.edu/cdm/archivalhome/collection/ead>), *Content, Context, and Capacity: A Collaborative Large-Scale Digitization Project on the Long Civil Rights Movement in North Carolina* (<http://www2.trln.org/ccp/index.htm>), and the John Muir Collections of the Holt-Atherton Special Collections Library at the University of the Pacific (<http://www.pacific.edu/Library/Find/Holt-Atherton-Special-Collections/John-Muir-Papers/John-Muir-Collections-.html>).

While “boutique” digitization projects usually involve the creation of a dedicated portal for access to digitized materials, large-scale digitization initiatives typically make use of online finding aids as a portal for access (Chapman and Leonard, 2013). In the article “Enduring Access to Special Collections: Challenges and Opportunities for Large-Scale Digitization Initiatives,” (2010) Rieger asserts that when using digitized archival materials, finding aids are essential to locating collections and understanding the composition of collections. Linking digital folders to their place in online finding aids also addresses scholars need to examine materials in their original context and maintain the provenance of documents as a body of related material (Greene, 2010; Sexton, 2002).

The use of online finding aids as the portal for large-scale digitization initiatives directly effects the nature of metadata associated with digitized materials. Metadata available to users for digitized materials is generally the same as descriptive information in the finding aid. The essential elements of archival metadata need to facilitate the discovery and access of archival materials (Rieger, 2010). It is thus imperative that archivists describe collections to support discovery and access. Rather than describing every single digitized item, librarians and archivists assert that if series and files are well described, they will provide sufficient information to direct users in their search (Greene, 2010; Sutton, 2012). What is more, attempting to provide item-level metadata for digitized material has been shown to slow and prevent progress in digitization projects (Greene, 2010). In the case of the John Muir Papers digitization project at the University of the Pacific, only pre-existing descriptions were used in the metadata, an approach which has garnered positive feedback for the ease of searching (Sutton, 2012). While it is true that not all finding aids are created equal, and may not provide sufficient description

to generate aggregate-level metadata, archivists are encouraged to begin thinking of large-scale digitization as a program rather than a project, and to embed such components of description and practice in the organizational structure (Rieger, 2010).

Archival Users

While current literature explores the theory behind large-scale digitization and best practices for conducting a large-scale digitization initiative, it lacks a thorough discussion of the usability of the product of large-scale digitization initiatives. Archivists at the University of Alabama recently conducted a usability test to evaluate searching for known items in the Septimus D. Cabaniss Papers digitization project (DeRidder, 2012). Results were inconclusive, likely due to the nature of the user group, as the majority were classified as novice. There is also a lack of inquiry regarding user satisfaction with the presentation and functionalities of online interfaces employed in large-scale digitization projects. Recognizing the needs of users of archives is central to facilitating a wider use of historical information in many facets of society (Conway, 1986). The literature broadly defines archival users as people who seek information, although the type of users studied varies widely, and may include undergraduate students, graduate students, and experienced historians and researchers (Conway, 1986; Murugan, 2011). The majority of scholars also agree that user studies should inform the design of archival systems. Jimerson argues that archivists need to identify their clientele and the design services that will suit their needs (Jimerson, 2003). Significantly, he highlights the assertion that archivists should not only identify users, but also understand their users and how they use the collections, a point which many scholars fail to address. He also argues, however, that users of archives “seek solutions to their information needs, not specific items,” a point of

contention among some archivists. Most archivists and scholars agree, however, that their primary audience should inform their design choices, and not the infrequent or single-visit patron searching for one specific solitary item (Proffitt, 2006).

Conversations among archivists and researchers have shed some light on how users of archives operate. In Duff and Johnson's 2002 study, "Accidentally Found On Purpose: Information-Seeking Behavior of Historians in Archives," they conducted semi-structured interviews with ten midcareer historians in an effort to investigate how they perform research and use archives. They identified four main activities: orienting to the archives and archival systems, seeking known material, building contextual knowledge, and identifying relative material. Similarly, archivists working in the Southern Historical Collection conducted interviews with a small group of scholars of the American South to inform their design of a large-scale digitization program (Southern Historical Collection, 2009). The group of scholars expressed a desire to have whole collections digitized at the aggregate level, as opposed to single items deemed to be of interest or importance by the archivist.

Methodology

This study is a qualitative content analysis of special collection libraries and archives identified as practicing large-scale digitization of archival materials. The purpose of the study is to examine how these institutions present the product of large-scale digitization projects to users on the Internet. Qualitative content analysis is “the study of recorded human communications,” as they appear in books, newspapers, emails, interviews, and in this case, web pages (Babbie 2010, 333). Qualitative content analysis is an appropriate method for this study because it allows the researcher to “examine meaning, themes, and patterns” that may be present in a text, as well as incorporating the specific context of the texts in the analysis (Zhang and Wildemuth 2009, 308). Qualitative analysis thus differs from quantitative analysis in that results are descriptive, and analysis focuses on observable themes and trends, rather than counting and statistical analysis (Babbie 2010, 340).

Institution	Web Address
Archives of American Art, Smithsonian Institution	www.aaa.si.edu/collections/online
Wilson Special Collections Library, University of North Carolina at Chapel Hill	http://www2.lib.unc.edu/wilson/
Princeton University Library	http://findingaids.princeton.edu/

University of Alabama Libraries	http://acumen.lib.ua.edu/
Colorado State University Libraries	http://lib.colostate.edu/digital-collections/
University of Maryland Libraries Digital Collections	http://digital.lib.umd.edu/
University of the Pacific Holt-Atherton Special Collections	http://www.pacific.edu/Library/Find/Holt-Atherton-Special-Collections.html
University of Wisconsin Digital Collections	http://uwdc.library.wisc.edu/collections
John F. Kennedy Presidential Library and Museum	http://www.jfklibrary.org/Research/Search-Our-Collections/Browse-Digital-Collections.aspx
Duke University Libraries	http://library.duke.edu/digitalcollections/

Table 1

The unit of analysis for the study was gathered through an analysis of the available literature in academic journals regarding large-scale digitization of archival materials. A list of institutions highlighted in the literature as practitioners of large-scale digitization was compiled, and a group of ten institutions were randomly selected for analysis. The selected institutions and websites used in the analysis can be seen in Table 1. An advantage of qualitative content analysis is that it allows for the purposeful selection of a unit of analysis in order to inform the research questions being investigated (Zhang and Wildemuth 2009, 309). Limiting the unit of analysis to institutions identified in the literature also served to eliminate researcher bias in the selection based on personal understanding of the definition of large-scale digitization.

Category	Details
Discovery and Access	Is there a link to digitized material on the library homepage? Is there a list of all collections with digitized content somewhere on the website? Does the institution use online finding aids as the platform for access to digitized materials? Is there a notification at the top of the finding aid alerting users to the existence of digitized content?
Materials	Has the institution digitized an entire collection? Has the institution digitized an entire series in a collection? Do digitized collections contain textual materials? Do digitized collections contain photographic materials? Do digitized collections contain audio/visual materials?
Metadata	Do digitized materials have item-level metadata? Do digitized materials have aggregate-level metadata? Can users contribute metadata to digitized materials?
Functionalities	Can users search across collections with user-generated keyword terms? Can the user adjust the size of an image? Can the user effectively navigate among images in a container? Can the user download digitized material? Can the user perform full-text searching on digitized documents? Are transcriptions of digitized documents available?

Table 2

After compiling the unit of analysis for the study, a codebook was written to examine the information and functionalities available to researchers when using online materials presented through large-scale digitization. Variables or markers, based on issues addressed in scholarly articles concerning large-scale digitization, were coded into the codebook found in Table 2. Variables were then divided into the following categories: discovery and access, materials, metadata, and functionalities. The coding scheme was analyzed and adjusted throughout the study to ensure consistency. In further efforts to ensure consistency, the researcher coded the websites over two consecutive days, using the same computer and internet browser. The results were recorded in an Excel spreadsheet and checked for appropriate consistency, as extremely inconsistent results could suggest an error in the coding scheme.

Results and Discussion

Discovery and Access

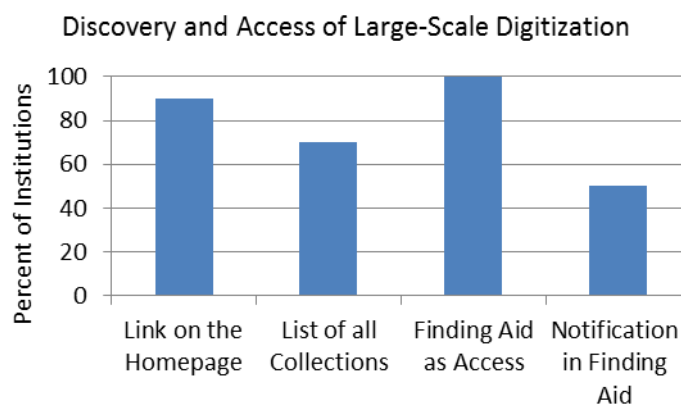


Figure 1

In order to increase access to special collections, it is important for libraries and archives to call attention to digitized material and ensure users are aware of online availability of materials. When surveying the websites of selected institutions, ninety percent display links to digitized material or collections on the library homepage. Institutions either provide a link to digitized material in the top navigation menu under “Research” or “Collections,” or include a post on the homepage directing users to digital collections and material. Many of the institutions sampled employ both methods to call attention to digitized material on the homepage. The University of Alabama website directs users to the “Digital Archive.” Of the institutions surveyed, Princeton University alone does not have a direct link to digitized materials on the library homepage. The homepage for Princeton University library finding aids contains a search box, allowing

users to search the content of all finding aids by keyword and optional date. Users can also browse archival materials by topic, names, and collections. There is not, however, a way to navigate directly to all collections with digitized content. It is left to the user to locate available digitized materials.

Another way to promote access to digitized special collections and attract potential users is to list all collections that contain digitized content. While it was expected that institutions with links to digitized material on the homepage would provide a list of all collections with digitized content, the results are somewhat different. Seventy percent of institutions surveyed do provide a list of all completely digitized collections and collections with digitized content somewhere on the website. A link to the list most often appears on the homepage for digital collections. The navigation link to “Digital Archive” on the University of Alabama homepage takes users directly to the list of collections with digitized content. List items are usually hyperlinks to collection finding aids. Institutions with comparatively fewer digital collections, such as the University of Maryland, are able to list all digital collections on a single HTML page. Other institutions with more extensive digital holdings, like the University of Wisconsin, allow users to browse digital collections through an alphabetical directory. Thirty percent of institutions surveyed do not provide a complete list of digitized collections. Princeton University, in addition to not having a link to digitized material on the library homepage, does not provide a comprehensive list of collections with digitized content. Colorado State University and Duke University provide a complete listing of digitized items, such as individual scans and documents, but do not provide a list of the collections from whence they came. In Duke University’s interface, item-level metadata identifies the parent

collection, and users can filter results by specific collections with faceted search terms in the navigation menu. The absence of a list of collections containing digitized material may ultimately be a result of the use of item-level metadata for digitized materials as opposed to aggregate-level metadata.

Online finding aids are widely accepted as an effective portal to digitized materials produced in large-scale digitization projects. Institutions frequently use online finding aids as the platform for discovery and access of digitized materials. It is important to note that all institutions surveyed had online finding aids, regardless of digitized content. Smaller institutions may lack the technical support to produce online finding aids. One hundred percent of institutions surveyed provided some access to digitized material through online finding aids. Of the libraries and archives sampled, only the Archives of American Art and Wilson Library exclusively use finding aids as the portal to digitized material. Interestingly, at the Archives of American Art, collections without digitized content appear to have only minimal description, while at Wilson Library, there does not appear to be a relation between digitized content and the level or richness of archival description. The remainder of institutions surveyed provide access to digitized materials through a combination of online finding aids and online exhibits. Of these institutions, it is more common that digitized personal and family papers are accessed through finding aids, while artificially assembled collections are more often displayed in online exhibits or dedicated portals. In the case of the University of Maryland, only one digital collection links to a finding aid. The rest of the digital collections are arranged as online exhibits, where users can locate items through keyword or faceted search and view results in a list. In some cases, such as the JFK Presidential Library website, users have

the option to view both a list of digitized content and the collection finding aid which links to digitized content. The collection title, however, is a hyperlink to the list form, while a smaller link to the collection finding aid is below the title. While these libraries and archives offer users both methods of discovering digitized content, the more prominent placement of links to lists of all digitized items suggests that users are encouraged to use this portal before entering the finding aid.

While all institutions in the sample provide some level of access to digitized materials through online finding aids, it is not always easy to determine if collections do contain digitized content. Fifty percent of library and archive websites surveyed include some sort of notification at the top of finding aids to alert the user to the presence of digitized material. The Archives of American Art includes a statement at the top of finding aids containing digitized material, explaining that the collection has been digitized and giving an exact number of scans associated with the collection. The Wilson Library finding aids contain a purple box at the top, stating that part or all of the collection has been digitized. Thumbnail images of digitized materials appear at the top of finding aids in the University of Alabama Library. In addition, the University of Alabama identifies collections with digitized content in the browse list with a special icon. The University of the Pacific uses the same icon to signal digitized content across multiple levels of content and description. A small, eye-shaped icon appears next to collections with digitized content in the browse list, at the top of finding aids with digitized content, at the top of series with digitized content, and at the item-level within the container list. Duke University Library includes a “digitized” icon next to collections

in the browse list and a banner highlighting digitized content at the top of finding aids.

Users also have the ability to limit a finding aid view to only digitized content.

Of the institutions that do not explicitly call out digitized content at the top of finding aids, most include icons in the browse list or within the container list in finding aids that highlight digitized materials. In Princeton University finding aids, users are not alerted to the existence of digitized content before navigating to a specific folder.

Colorado State University Library finding aids contain links to digitized content only at the item level. The JFK Presidential Library website places “digitized” icons next to collections in the browse list, but there is no indication at the top of the finding aid that the collection contains digitized materials. It is worth noting that the JFK Presidential Library website primary directs users to the list view of digitized content, and not to online finding aids. Institutions that primarily direct users of digitized materials to online finding aids are more likely to call attention to the existence of digitized content at the top of the finding aid.

Digitized Material

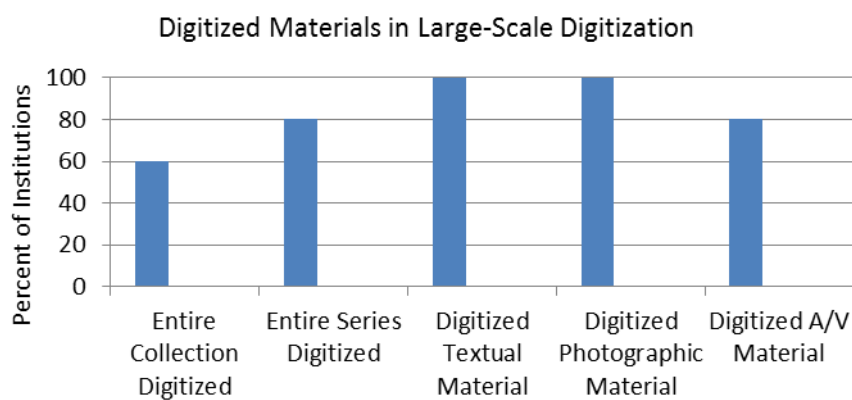


Figure 2

The available literature indicates that large-scale digitization is not defined by the number of items scanned, but by whether materials are scanned at the aggregate level. Aggregate level could mean collection, series, box, or folder. The goal of most large-scale digitization projects is to digitize entire collections or whole series within collections. Sixty percent of institutions surveyed provide access to entirely digitized collections online. The Archives of American Art, the Wilson Library, and the JFK presidential library explicitly state on their websites that some collections have been fully digitized. It is evident from examining finding aids from the University of Alabama, the University of Maryland, and Duke University that entire collections have been digitized. A higher percentage of institutions surveyed have digitized entire series, if not entire collections. Eighty percent of institutions have digitized materials at the series level, while twenty percent have digitized entire series, but not collections. Of the four institutions coded as not having digitized entire collections, it was impossible for the researcher to identify fully digitized collections. The institution websites did not explicitly state that entire collections had been digitized, nor was it possible to determine from examining the finding aids if collections had been digitized in their entirety. The majority of digital collections at the University of Wisconsin are artificial collections that have been assembled for online exhibits. It is not possible to determine the parent collection of most digitized items, and therefore impossible to identify fully digitized collections.

The goal of large-scale digitization is for archivists and librarians to select materials at the aggregate or container level, as opposed to identifying individual items for digitization. The goal of this approach is to recreate the experience of using a physical

collection in the library and allow researchers to draw contextual information through analyzing all items in a container. In principle, large-scale digitization requires that all materials in a container be digitized, regardless of format. Among the institutions surveyed, one hundred percent have digitized both textual and photographic material, while an admirable eighty percent have digitized some type of audio/visual material. Of the institutions that do not provide online access to digitized audio/visual materials, the Archives of American Art is currently conducting an ongoing project funded by the Council on Library and Information Resources “Hidden Collection” grant program to digitize hidden audio/visual material.² Use copies of digitized materials, however, are only available to researchers in the archive reading room as they become available, and are not accessible online. Digitization practices, scanning techniques, workflows, and access methods for paper-based archival materials have been well documented. The varied formats of audio/visual materials present a range of new challenges for digitization, including the need for specialized equipment, technicians with special training, and the capability to serve audio and video files on the web. In light of these limitations, it is promising for the future of audio/visual digitization that eighty percent of institutions surveyed provide online access to audio and video.

² For more information regarding the grant-funded project “Uncovering Hidden Audiovisual Media Documenting Postmodern Art,” see <http://www.aaa.si.edu/collections/documentation/av>.

Metadata

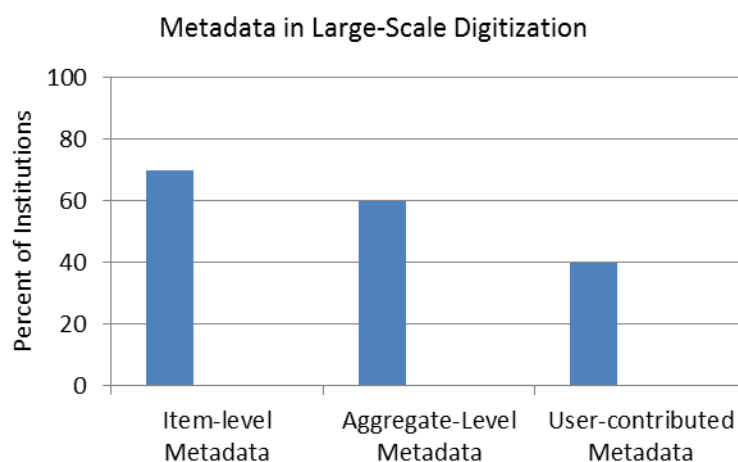


Figure 3

The selection process in large-scale digitization is directly related to the level of metadata associated with digitized material. Because materials in large-scale digitization projects are not individually selected, it is difficult and time consuming to assign item-level metadata to digitized collections. Large-scale digitization essentially trades enhanced metadata for larger amounts of digitized material. In spite of the difficulties in providing item-level metadata with large-scale digitization, seventy percent of institutions surveyed provide some item-level metadata for digitized content, although not necessarily for all digitized materials. In Princeton University finding aids, some digitized content has item-level metadata because materials are described at the item level. Duke University provides item-level metadata for some digitized content, but not as part of large-scale digitization. Materials with item-level metadata are most often part of an online digital exhibit or artificial collection. Sixty percent of institutions surveyed include aggregate or container-level metadata with digitized materials. The Archives of American Art includes both the folder number and title with digitized material, as well as folder date ranges where available. Princeton University, the JFK Presidential Library,

and Duke University also provide folder-level metadata with large-scale digitization. The Wilson Library provides container level metadata for digitized material, but does not limit the definition of container to a folder. Container types include folder, box, photograph, photograph album, oversize paper, digital file, etc. Series and collection information is also included where available. Of the forty percent of institutions surveyed that do not provide users with aggregate-level metadata, all include item-level metadata with digitized material. The Archives of American Art, the Wilson Library, and the JFK Presidential Library provide exclusively aggregate-level metadata. The researcher found no evidence of item-level metadata assigned to digitized materials.

Because many large-scale digitization projects provide aggregate-level metadata for digitized materials, it is often left to the researcher to identify people, places, or events described or depicted in individual items. In an attempt to crowd source generating item-level metadata for digitized collections, librarians and archivists have enabled online interfaces to allow users to tag or comment on digitized items. While some institutions may be limited by technical capabilities, forty percent of institutions surveyed provided some method for users to contribute metadata. The Wilson Library and the University of the Pacific have enabled commenting and tagging capabilities in the CONTENTdm interface, allowing users to contribute metadata at the item level. Duke University has enabled a comment box on certain digitized items in online exhibits, but users cannot contribute metadata for items accessed through online finding aids. Princeton University provides users with a comment box at the aggregate level, but users cannot assign comments to an individual digitized image.

Functionalities

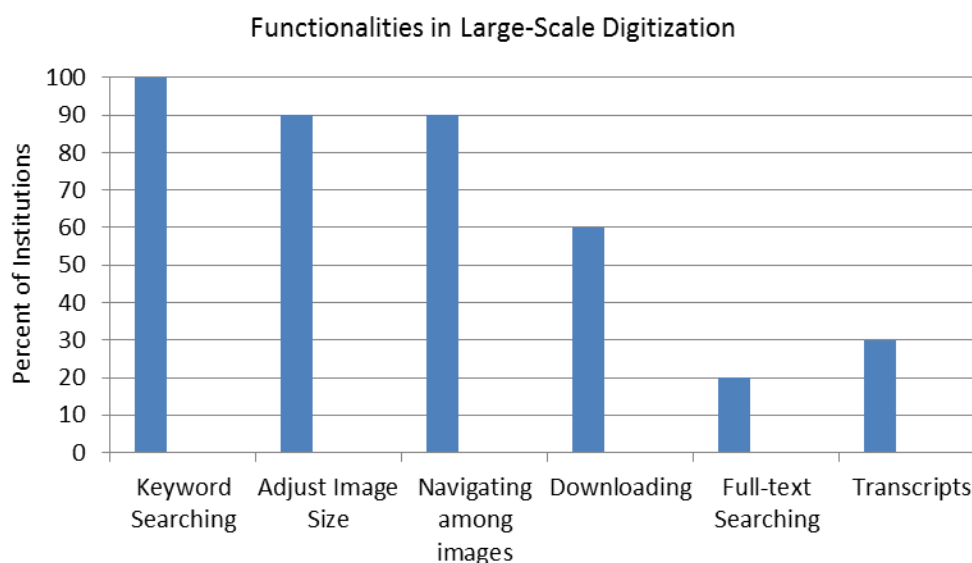


Figure 4

While many institutions hold that the goal of large-scale digitization is to recreate online the experience of performing special collection research in a physical reading room, the web presents extensive possibilities for searching and manipulating documents that are not possible in the physical realm. One hundred percent of institutions surveyed allowed users to search across collection descriptions with user-generated keywords. All institutions displayed a search box on the homepage for digital collections. The University of the Pacific allows users to perform keyword searches at different levels for individual collections. For example, in the John Muir Correspondence, users can perform keyword searches in the following categories: “Full-Text Transcriptions,” “Correspondence From,” “Correspondence To,” “Original Date,” and “Owning Institution.” Few of the library and archives surveyed, however, allow keyword searching across only digitized material, rather than returning results from across the website or catalog.

A common complaint heard in special collections reading rooms is that text in hand-written manuscripts is often too small or difficult to read. Ninety percent of institutions surveyed allowed users some method of changing the viewing size of digitized materials. The Archives of American Art provides a scroll bar to zoom in and out within an image. Users of the Wilson Library's digitized materials can toggle sizes in the light box view, and zoom in and out within an image when viewed in CONTENTdm. The University of Alabama allows users to zoom and fit the image to the screen. Users of the University of Maryland's digitized materials can adjust the size of an image, but only on the download page. Princeton University, the one institution that does not allow users to change the size of the image, allows users to rotate an image.

Special collections researchers have expressed the need to easily navigate from one image to the next in a digitized container, similar to flipping through a folder of documents in the reading room. Ninety percent of institutions surveyed provided an effective way to navigate between scans in a container. The primary navigation methods are arrows keys to click through a container, or thumbnail views of the entire container in a fixed header or sidebar. The Archives of American Art includes a sidebar of thumbnail views of all images in a digitized container. Users can scroll through the images and select individual scans to view. The Wilson Library allows users to move backwards and forwards within a container, and also provides a slideshow option. The University of Alabama displays a thumbnail ribbon, or "film strip" as a header in the viewing frame for digitized materials.

An advantage to digitizing special collection material is that items are scanned once, rather than being photocopied repeatedly for multiple users. Users can save copies

of digitized material for their personal use. It is not, however, standard practice for institutions to allow users to download all or any of their digitized collections. Sixty percent of institutions surveyed provided some way for users to download digitized material. The Archives of American Art does not explicitly allow users to download materials, but users can save materials from the print screen. The Wilson Library directs users to a “Downloadable Image” of the highest resolution available. Users can then save the image to a specific location. Both Princeton University and the University of Wisconsin allow users to download a PDF file of the entire container, as opposed to only single images. The University of Alabama, the University of Maryland, the University of the Pacific, and the JFK Presidential Library do not allow users to download digitized materials. Possible reasons may include copyright restrictions, use restrictions, agreements with donors, or sensitive information.

While manual transcription is feasible in small, item-level digitization projects, the scale of materials scanned in large-scale digitization projects essentially prohibits manual transcription. It is not within the budgets of most institutions to employ staff in manually transcribing thousands of pages of documents. A popular trend in large-scale digitization is the use of Optical Character Recognition (OCR) software to produce text files of digitized documents. While institutions have seen varied results in the accuracy of transcriptions produced through OCR software, a few have begun offering full-text searching capabilities with digitized archival materials. Twenty percent of institutions surveyed provide some degree of full-text searching capabilities for their digitized collections, while thirty percent provide transcriptions for some digitized material. No institution provides transcriptions or full-text searching for all digitized content. The

University of Wisconsin provides full-text searching capabilities in the US Foreign Relations collection. The University of Maryland and Duke University provide transcripts of digitized materials in select cases, such as the American Sheet Music collection at Duke. These cases are, however, limited to smaller digital exhibits or collections. The University of the Pacific allows full-text searching of digitized documents in the John Muir Correspondence, in addition to providing transcriptions of materials. The collection, however, was digitized as part of a grant-funded project, with the goal of producing searchable transcripts. This level of transcription is likely not sustainable for a long-term digitization initiative.

Conclusion

The purpose of this study was to examine how archives and special collection libraries present and provide access to materials digitized in a large-scale digitization initiative. The goal was to identify both trends and variances in the tools and capabilities available to users of digitized archival materials. A content analysis was performed on these tools and capabilities by examining the websites and online interfaces of institutions identified as implementing large-scale digitization initiatives. The goal of the study was that the results would highlight strengths and weaknesses in how institutions allow users to interact with online digitized material.

Through analysis of the websites of the selected special collection libraries and archives, this study has shown that the institutions effectively call attention to the existence of digitized collections and materials. This is an important measure, as many potential users will be reluctant or unable to visit the physical repository. Casual users are also likely to leave the site if they do not quickly locate digitized content. In the digital age, users expect to find digitized materials, and it is important that special collection websites effectively direct them to the content. The analysis shows, however, a lack of consistency in how institutions alert researchers to the existence of digitized material in a particular collection once they are in the finding aid. Researchers may not arrive at a finding aid through the homepage for digitized collections, and it is important that they know the material they are searching for may be digitized.

While the literature clearly states that the goal of large-scale digitization initiatives is not to produce item-level metadata, analysis of the websites revealed an inconsistency in this practice. It was hypothesized at the beginning of this study that most, if not all institutions would provide aggregate-level metadata for digitized materials. The high percentage of institutions implementing item-level metadata indicates continuing inconsistencies and confusion regarding the definition and characteristics of large-scale digitization. It is possible that enhanced metadata was added after the initial digitization effort. Clarity and consistency in what users can expect from large-scale digitization will go a long way in improving the user experience across institutional interfaces.

While the majority of institutions have made a successful effort to digitize all material formats encountered in large-scale digitization in spite of difficulties, most institutions fall short in gathering user-contributed metadata and allowing full-text searching. Many archivists are wary of allowing unknown users to contribute metadata that may be seen as authoritative by other users. A potential solution for reluctant repositories is to gather a group of “super users,” or experts in a particular collection or field to provide enhanced metadata for a defined set of items. This approach may help ease archivists into adopting user-contributed metadata. In terms of full-text search capabilities of digitized material, archivists are unfortunately limited by the quality and functionality of available OCR software. This is an issue that archivists and librarians must continue to explore, as it will further increase the discoverability of digitized material and enhance the user experience.

One of the main lessons learned in conducting this analysis is that inconsistencies in practices and interfaces for using digitized materials contribute to a negative user experience. Mastering the website, interface, and tools of one library or archive does not guarantee ease of use of another institution's website. While this content analysis identifies several trends in how users can interact with materials produced in large-scale digitization initiatives, further study is required to determine which functions and designs best serve the needs of archival users. A usability study of several different interfaces for large-scale digitization is a logical next step. Improving the interaction users have with digitized archival materials will increase the chances of new users returning, thus widening the scope of archival users and promoting access to our historical and cultural treasures.

Bibliography

- Babbie, Earl R. (2010). *The Practice of Social Research*, 12th ed. Belmont, CA: Thomson Wadsworth.
- Cenyowa, Klaus. (2009). Mass Digitization for Research and Study: the Digitization Strategy of the Bavarian State Library. *IFLA Journal* 35 (1): 17-24.
- Chapman, Joyce and Samantha Leonard. (2013). Cost and Benefit of Quality Control Visual Checks in Large-Scale Digitization of Archival Manuscripts. *Library Hi Tech* 33 (3): 406-418.
- Conway, Paul. (1986). Facts and Frameworks: An Approach to Studying the Users of Archives. *The American Archivist* 49: 393-407.
- DeRidder, Jody L., Amanda Axley Presnell, and Kevin W. Walker. (2012). Leveraging Encoded Archival Description for Access to Digital Content: A Cost and Usability Analysis. *The American Archivist* 75: 143-170.
- Greene, Mark A. (2010). MPLP: It's Not Just for Processing Anymore. *The American Archivist* 73: 175-203.
- Gueguen, Gretchen. (2010). Digitized Special Collections and Multiple User Groups. *Journal of Archival Organization* 8 (2): 96-109.
- Jimerson, Randall C. (2003). Reference, Access, and Use. *OCLC Systems & Services* 19 (1): 13-16.
- Murugan, V. Senthurvel. (2011). User Studies and User Education. *International Journal of Library and Information Science* 4 (1/2): 87-110.
- Profitt, Merrilee. (2006). How and Why of User Studies: RLG's RedLightGreenLight as a Case Study. *Journal of Archival Organization* 4 (1/2): 87-110.
- Southern Historical Collection, University Library, University of North Carolina at Chapel Hill. (2009). Extending the Reach of Southern Sources: Proceeding the Large-Scale Digitization of Manuscript Collections. Final grant report for the Andrew W. Mellon Foundation. Retrieved 2 December 2013 from http://www.lib.unc.edu/mss/archivalmassdigitization/download/extending_the_reach.pdf.

- Rieger, Oya Y. (2010). Enduring Access to Special Collections: Challenges and Opportunities for Large-Scale Digitization Initiatives. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 11: 11-22.
- Sexton, Anna and Christ Turner. (2002). Expanding the Role of EAD: Providing Adequate Metadata for Digital As Well As Original Archival Documents. *VINE* 32 (4): 71-80.
- Smith, Kevin L. (June 2012). Copyright Risk Management: Principles and Strategies for Large-Scale Digitization Projects in Special Collections. *Research Library Issues: A Quarterly Report from ARL, CNI, and SPARC*, (279): 17-23.
- Sutton, Shan C. (2012). Balancing Boutique-Level Quality and Large-Scale Production: The Impact of “More Product, Less Process” on Digitization in Archives and Special Collections. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 13: 50-63.
- Weintraub, J and M. A. Wisner. (October 2008). Mass Digitization at Yale University Library: Exposing the Treasures in Our Stacks. *Computers in Libraries* 28 (9): 10, 12-6.
- Zhang, Yan and Barbara M. Wildemuth. (2009). Qualitative Analysis of Content. From *Applications of Social Research Methods to Questions in Information and Library Science*. Westport, Conn.: Libraries Unlimited: 308-319.