

**Shopping the Online Archives Megastore: A  
Content Analysis of Special Collections  
Libraries and Archives Websites Produced  
Through Large-Scale Digitization**  
by Amanda L. Loeb

**Abstract**

This study is a content analysis examining websites of special collections 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 conducting large-scale digitization. The results demonstrate success in highlighting digitized materials on the resultant 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.

**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 collections are no longer sufficient to meet the needs of users who want to shop the online archives megastore. Archivists and librarians have responded by implementing large-scale digitization of materials, providing online access to large volumes of materials. 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 collections libraries and archives supporting large-scale digitization initiatives in order to better understand how users can interact with resources produced through this 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.<sup>1</sup> 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 similar projects, the Bavarian State Library brought more than 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. The goal of large-scale digitization is generally to provide 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.<sup>2</sup>

The current literature identifies selection of materials as a defining component of a large-scale digitization initiative. In these projects, the archivist performs minimal selection of materials, 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 a greater extent of materials, whereas the time and resources required for boutique digitization often limits the scope of projects.<sup>3</sup> Archivists hold that this approach accomplishes the request of scholars and archival users to preserve the context of archival materials in the digital environment.<sup>4</sup> Large-scale digitization initiatives also support users' expectations of accessing large quantities of information via the web.<sup>5</sup> 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 materials.<sup>6</sup>

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.<sup>7</sup> In the article "Enduring Access to Special Collections: Challenges and Opportunities for Large-Scale Digitization Initiatives," Oya Rieger asserts that when using digitized archival materials, finding aids are essential to locating collections and understanding the composition of collections.<sup>8</sup> Linking digital folders to their place in online finding aids also addresses scholars' needs to examine materials in their original context and maintain the provenance of documents as a body of related materials.<sup>9</sup>

The use of online finding aids as the portal for large-scale digitization initiatives directly affects 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.<sup>10</sup> 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. What is more, attempting to provide item-level metadata for digitized materials has been shown to slow and prevent progress in digitization projects. 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 that garnered positive feedback for the ease of searching.<sup>11</sup> While it is true that not all finding aids are created equal, and they 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.<sup>12</sup>

### *Archival Users*

While current literature explores the theory behind large-scale digitization and best practices for conducting such a program, 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.<sup>13</sup> Results were inconclusive, likely due to the nature of the user group, as the majority of users 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.<sup>14</sup> 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.<sup>15</sup> The majority of scholars also agree that user studies should inform the design of archival systems. Randall Jimerson argues that archivists need to identify their clientele and design services that will suit their needs.<sup>16</sup> 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 that 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.<sup>17</sup>

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 mid-career 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 materials, building contextual knowledge, and identifying relative materials.<sup>18</sup> Similarly, archivists working in the Southern Historical Collection at The University of North Carolina at Chapel Hill conducted interviews with a small group of scholars of the American

South to inform their design of a large-scale digitization program.<sup>19</sup> 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 collections 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.<sup>20</sup> 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.<sup>21</sup> 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.<sup>22</sup>

The unit of analysis for the study was gathered through an examination 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 this analysis can be found 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.<sup>23</sup> Limiting the

**Table 1: Institutions and Corresponding Websites Analyzed**

<b>Institution</b>	<b>Web Address</b>
Archives of American Art, Smithsonian Institution	<a href="http://www.aaa.si.edu/collections/online">www.aaa.si.edu/collections/online</a>
Wilson Special Collections Library, University of North Carolina at Chapel Hill	<a href="http://www2.lib.unc.edu/wilson/">http://www2.lib.unc.edu/wilson/</a>
Princeton University Library	<a href="http://findingaids.princeton.edu/">http://findingaids.princeton.edu/</a>
University of Alabama Libraries	<a href="http://acumen.lib.ua.edu/">http://acumen.lib.ua.edu/</a>
Colorado State University Libraries	<a href="http://lib.colostate.edu/digital-collections/">http://lib.colostate.edu/digital-collections/</a>
University of Maryland Libraries Digital Collections	<a href="http://digital.lib.umd.edu/">http://digital.lib.umd.edu/</a>
University of the Pacific Holt-Atherton Special Collections	<a href="http://www.pacific.edu/Library/Find/Holt-Atherton-Special-Collections.html">http://www.pacific.edu/Library/Find/Holt-Atherton-Special-Collections.html</a>
University of Wisconsin Digital Collections	<a href="http://uwdc.library.wisc.edu/collections">http://uwdc.library.wisc.edu/collections</a>
John F. Kennedy Presidential Library and Museum	<a href="http://www.jfklibrary.org/Research/Search-Our-Collections/Browse-Digital-Collections.aspx">http://www.jfklibrary.org/Research/Search-Our-Collections/Browse-Digital-Collections.aspx</a>
Duke University Libraries	<a href="http://library.duke.edu/digitalcollections/">http://library.duke.edu/digitalcollections/</a>



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.

After compiling the group of institutions and websites for the study, a set of analysis criteria, or codebook, was created to examine the information and functionalities available to researchers when using online materials presented through large-scale digitization. Variables based on issues addressed in scholarly articles concerning large-scale digitization were included in the codebook, found in Table 2. Variables were then divided into the following categories: discovery and access, materials, metadata, and functionalities. The analysis variables were evaluated and adjusted throughout the study to ensure consistency. In further efforts to ensure consistency, the researcher analyzed the websites over two consecutive days, using the same computer and Internet browser. The results were recorded in

**Table 2: Evaluation Algorithm**

<b>Category</b>	<b>Detailed Criteria</b>
Discovery and Access	<p>Is there a link to digitized materials on the library home page?</p> <p>Is there a list of all collections with digitized content somewhere on the website?</p> <p>Does the institution use online finding aids as the platform for access to digitized materials?</p> <p>Is there a notification at the top of the finding aid alerting users to the existence of digitized content?</p>

**Table 2: Evaluation Algorithm (continued)**

<b>Category</b>	<b>Detailed Criteria</b>
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 audiovisual 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 materials? Can the user perform full-text searching on digitized documents? Are transcriptions of digitized documents available?

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*

In order to increase access to special collections, it is important for libraries and archives to call attention to digitized materials and ensure that users are aware of online availability of materials. When surveying the websites of selected institutions, 90 percent display links to digitized materials or collections on the library home page. These institutions either provide a link to digitized materials in the top navigation menu under “Research” or “Collections,” or include a post on the home page directing users to digital collections and materials. Many of the institutions sampled employ both methods to call attention to digitized materials on the home page. 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 home page. The home page 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 materials on the home

page 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 home page for digital collections. The navigation link to “Digital Archive” on the University of Alabama home page takes users directly to the list of collections with digitized content. The items listed 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, which does not have a link to digitized materials on the library home page, also does not provide a comprehensive list of collections with digitized content. Colorado State University and Duke University each 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 materials 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. It is important to note that all institutions surveyed had online finding aids, regardless of digitized content. Smaller institutions, however, may lack the technical support to produce online finding aids. One hundred percent of institutions surveyed provided some access to digitized materials through online finding aids. Of the libraries and archives sampled, only the Archives of American Art and the Wilson Library exclusively use finding aids as the portal to digitized materials. Interestingly, at the Archives of American Art, collections without digitized content appear to have only minimal description, while at the Wilson Library, there does not appear to be a relation between digitized content and the level or richness of archival description. The remaining institutions 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 that 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 materials. The Archives of American Art includes a statement at the top of finding aids containing digitized materials 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 for the University of Alabama Libraries. 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 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 Libraries 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 primarily 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 Materials*

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. 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. While not explicitly stated, 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 there as well. 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 20 percent have digitized entire series but not

collections. Of the four institutions determined 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.

Large-scale digitization allows archivists and librarians to select materials at the aggregate or container level, as opposed to identifying individual items for digitization. One 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, 100 percent have digitized both textual and photographic materials, while an admirable 80 percent have digitized some type of audiovisual materials. Of the institutions that do not provide online access to these materials, the Archives of American Art is currently conducting an ongoing project funded by the Council on Library and Information Resources “Hidden Collections” grant program to digitize hidden audiovisual materials.<sup>24</sup> Use copies of digitized materials, however, are available to researchers in the archives’ reading room as they become available, but are not accessible online. While digitization



practices, scanning techniques, workflows, and access methods for paper-based archival materials have been well documented, the varied formats of audiovisual materials present a range of new challenges, 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 audiovisual digitization that 80 percent of institutions surveyed provide online access to audio and video files.

### *Metadata*

The selection process in large-scale digitization is directly related to the level of metadata associated with digitized materials. Because materials in large-scale digitization projects are not individually selected, it is difficult and time consuming to assign item-level metadata to these collections. Large-scale digitization essentially trades enhanced metadata for larger amounts of digitized materials. In spite of the difficulties in providing item-level metadata with large-scale digitization projects, 70 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 already described at the item level. Duke University also 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 number and title of folders containing digitized materials, 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 materials, 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 40 percent of institutions surveyed that do not provide users with aggregate-level metadata, all include item-level metadata with digitized materials. 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 crowdsource the generation of item-level metadata for digitized collections, librarians and archivists enabled online interfaces to allow users to tag or comment on digitized items. While some institutions may be limited by technical capabilities, 40 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*

While many institutions hold that the goal of large-scale digitization is to recreate online the experience of performing special collections 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 home page 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 materials 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 collections materials is that items are scanned once, rather than being photocopied repeatedly for multiple users. Users can save copies of digitized materials 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 some digitized materials. 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 volume 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 30 percent provide transcriptions for some digitized materials. No institution provides transcriptions or full-text searching for all digitized content. The University of Wisconsin provides full-text searching capabilities in the U.S. 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. However, the collection 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**

This study examined how archives and special collections libraries present and provide access to materials produced in a large-scale digitization initiative in order to identify both trends and variances in the tools and capabilities available to users of digitized archival materials. The results highlight strengths and weaknesses in how institutions allow users to interact with online digitized materials.

Through analysis of the websites of the selected special collections 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 materials in a particular collection once they are in the finding aid. Researchers may not arrive at a finding aid through the home

page for digitized collections, and it is important that they know the materials they are searching for might 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, 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 materials, 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 materials 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.

*Amanda L. Loeb is a processing archivist at the Louis Round Wilson Special Collections Library at the University of North Carolina at Chapel Hill. She is a recent graduate of the School of Information and Library Science at UNC with a concentration in archives and records management.*

## NOTES

1. Klaus Cenyowa, "Mass Digitization for Research and Study: the Digitization Strategy of the Bavarian State Library," *IFLA Journal* 35.1 (2009): 17-24, and J. 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 (2008): 10-16.
2. Joyce Chapman and Samantha Leonard, “Cost and Benefit of Quality Control Visual Checks in Large-Scale Digitization of Archival Manuscripts,” *Library Hi Tech* 33.3 (2013): 406-418 and Shan C. Sutton, “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 (2012): 50-63. Examples of large-scale digitization projects in special collections 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/ccc/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>).
  3. Sutton, 51.
  4. Oya Y. Rieger, “Enduring Access to Special Collections: Challenges and Opportunities for Large-Scale Digitization Initiatives,” *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 11 (2010): 11-22.
  5. Mark A. Greene, (2010). “MPLP: It’s Not Just for Processing Anymore,” *The American Archivist* 73.1 (2010): 175-203 and Gretchen Gueguen, “Digitized Special Collections and Multiple User Groups,” *Journal of Archival Organization* 8.2 (2010): 96-109.
  6. Kevin L. Smith, “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 (2012): 17-23.
  7. Chapman and Leonard, 406.
  8. Rieger, 16.

9. Greene, 194, and Anna Sexton and Christ Turner, "Expanding the Role of EAD: Providing Adequate Metadata for Digital As Well As Original Archival Documents," *VINE* 32.4 (2002): 71-80.
10. Rieger, 16.
11. Greene, 193-194, and Sutton, 52.
12. Rieger, 17.
13. Jody L. DeRidder, Amanda Axley Presnell, and Kevin W. Walker, "Leveraging Encoded Archival Description for Access to Digital Content: A Cost and Usability Analysis," *The American Archivist* 75 (2012): 143-170.
14. Paul Conway, "Facts and Frameworks: An Approach to Studying the Users of Archives," *The American Archivist* 49 (1986): 393-407.
15. V. Senthurvel Murugan, "User Studies and User Education," *International Journal of Library and Information Science* 4.1/2 (2011): 87-110.
16. Randall C. Jimerson, "Reference, Access, and Use," *OCLC Systems & Services* 19.1 (2003): 13-16.
17. Merrilee Profitt, "How and Why of User Studies: RLG's RedLightGreenLight as a Case Study," *Journal of Archival Organization* 4.1/2 (2006): 87-110.
18. Wendy M. Duff and Catherine A. Johnson, "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives," *The Library Quarterly: Information, Community, Policy* 72.4 (2002): 472-496.
19. 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, 2009), accessed 2 December 2013 from [http://www.lib.unc.edu/mss/archivalmassdigitization/download/extending\\_the\\_reach.pdf](http://www.lib.unc.edu/mss/archivalmassdigitization/download/extending_the_reach.pdf).
20. Earl R. Babbie, "The Practice of Social Research," (Belmont, CA: Thomson Wadsworth, 2010), 333.
21. Yan Zhang and Barbara M. Wildemuth, "Qualitative Analysis of Content," in *Applications of Social Research*

*Methods to Questions in Information and Library Science*, (Westport, Conn.: Libraries Unlimited, 2009): 308-319.

22. Babbie, 340.

23. Zhang and Wildemuth, 309.

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

Copyright of Journal for the Society of North Carolina Archivists is the property of Society of North Carolina Archivists and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

## **2013 Gene J. Williams Award Winner**

*The Gene J. Williams Award, presented annually by the Society of North Carolina Archivists, recognizes excellence for a paper on an archival topic written by a North Carolina graduate student for a graduate-level course. This award honors the late Gene J. Williams, archivist at the North Carolina Division of Archives and History and at East Carolina University and charter member of the Society of North Carolina Archivists.*

### **Bridging the Gap: Handling Born-Digital Records in Manuscript Repositories**

by Courtney Bailey

*This article is a version of Bailey's master's paper, which is available in full at <http://bit.ly/1x3JG16>*

#### **Abstract**

With a rapidly burgeoning volume of born-digital records, archives must determine how they can best bridge the gap between handling analog and born-digital records. In addition to analyzing existing case studies, the study presented in this paper used both a survey and semi-structured interviews with archivists to investigate whether and how manuscript repositories in particular are handling born-digital materials, in order to pinpoint problems and identify some practical solutions that suggest courses for further study and action. The difficulties of curating born-digital materials, especially when it comes to providing

access, are not unique, and the profession would be well-served by finding a space for collaboration to solve these thorny issues.

## **Introduction**

In his famous 1995 article for *Scientific American*, Jeff Rothenberg warns, “digital information lasts forever – or five years, whichever comes first.”<sup>1</sup> Elizabeth Dow echoes this point, recognizing that while analog materials can be accessioned and then processed at whatever point in the future is most convenient, electronic records have their own schedule that must be followed unless a repository is willing to risk losing them.<sup>2</sup> When Rothenberg expanded on this article four years later, he identifies four primary modes of loss of electronic records: physical decay of media; loss of information about the format, encoding, or compression of files; obsolescence of hardware; and unavailability of software with which to access the digital information.<sup>3</sup> While Rothenberg and others certainly called attention to the issue of electronic records in the 1990s, archivists have been collecting and curating machine-readable records at least since the 1970s. For example, the National Archives and Records Administration (NARA) received its first electronic records in 1969.<sup>4</sup> A few years before Rothenberg’s piece, Margaret Hedstrom suggests that “archivists need to be open to ‘radical thinking’ about the role of archives because successfully dealing with electronic records may demand a transformation of the basic purpose of archives and the methods archivists use.” She further identifies barriers to success, concluding that “fear of change, aversion to risk, a custodial mentality, and a failure to recognize electronic records as critical to the future success of archives” are all

significant impediments that reinforce the notion that the purview of archivists is limited to the paper realm.<sup>5</sup> Due to expanding varieties of digital objects and increasing debate over whether electronic records fit into the traditional paradigm of archival practice, a second wave of literature about born-digital materials was generated during the last decade of the twentieth century and the first decade of the twenty-first century. Setting aside disagreements over archival principles, one weakness of these writings from the perspective of manuscript repositories is their focus on the evidential value of business and government records without much acknowledgment of the role that personal papers play in the documentation of society. Writing in *Archives and Manuscripts* in 1994, Adrian Cunningham proclaims his article as “a first attempt at redressing this imbalance in the literature” regarding personal records versus government and organizational records.<sup>6</sup>

With a rapidly burgeoning volume of born-digital records, it behooves archives to determine how they can best bridge the gap between management of analog and born-digital records. In his foreword to the 2010 report *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*, Charles Henry estimates that 90 percent of the records currently being created are born digital.<sup>7</sup> With the huge number of born-digital materials being generated and the long history of their consideration in archival literature, one has to wonder why manuscript repositories have yet to reach a consensus about the best methods for handling born-digital collections. In *Electronic Records in the Manuscript Repository*, Elizabeth Dow offers one explanation. She suggests that electronic records “aren’t nearly as seductive as the old records” and refers to them as the “ugly babies of our

professional future” – reminding us that digital records will still need just as much care and attention as the traditional “pretty babies” of the archives, such as a letter by a famous author on monogrammed stationery.<sup>8</sup>

Laura Millar helps to further identify five key problem areas in the archiving of electronic records in her 2010 book *Archives: Principles and Practices*: “technological dependence and obsolescence; mutability; the potential loss of context; the effects of decentralized information management, security and privacy; and cost.”<sup>9</sup> Perhaps the most challenging aspect of this list for manuscript repositories is these factors reside outside their control. Software and hardware dependencies and/or obsolescence are set in motion by the record creators, therefore many repositories do not have involvement with or control over decisions such as whether to use open-source or interoperable software. The issues of context and decentralized control also depend upon record creators, while the collecting repository must be resigned to accept whatever amount of metadata is provided by the donor and whatever complications or losses of data occurred before materials crossed the archival threshold. While the repository is directly responsible for providing security that ensures records cannot be damaged or changed and personally identifiable information is not distributed to inappropriate parties – just as archives have done for centuries with analog records – the mechanisms for causing these problems are very different for electronic records; viruses or bit rot can affect the integrity of computer files. Finally, cost also remains a factor outside the control of repositories, with budgets often dependent on institutional funding or soft money from grants, the prices for technological components



set by market forces, and the staffing requirements very dependent on individual abilities and interests. All these uncertainties help to explain why many manuscript repositories have been somewhat reluctant to take up the mantle of preserving born-digital materials.

In 2010, OCLC Research surveyed the Association of Research Libraries (ARL), the Canadian Academic and Research Libraries, the Independent Research Libraries Association, the Oberlin Group, and the U.S. and Canadian members of the RLG Partnership. Jackie Dooley and Katherine Luce identify three actions that need to be taken by the special collections community to address born-digital archival materials: (1) “Define the characteristics of born-digital materials that warrant their *management as ‘special collections’*”; (2) “Define a reasonable set of *basic steps* for initiating an institutional program for responsibly managing born-digital archival materials ”; and (3) “Develop *use cases and cost models* for selection, management, and preservation of born-digital archival materials.” One of their survey questions revealed information about impediments to the management of born-digital materials, with the results showing lack of funding, lack of time for planning, lack of expertise, and lack of support within the institution as common roadblocks.<sup>10</sup> As already noted, lack of funding is often beyond the reach of a repository to solve, and lack of institutional support will require some intentional advocacy by archivists to reverse, but lack of time for planning and lack of expertise can immediately be addressed.

In 2012 Ricky Erway produced two reports for OCLC Research that contribute to the conversation about born-digital materials. In *You’ve Got to Walk Before You Can Run*, she suggests a simple three-step survey can begin

to address electronic records that are already in the collections of an archive: (1) find the physical media already in the repository, (2) count and describe these media, and (3) prioritize the further treatment of collections.<sup>11</sup> Simplifying the situation and identifying a place where repositories can begin attacking the problem of born-digital materials should enable more institutions to join this field. The second in this series of Demystifying Born Digital reports suggests the creation of SWAT sites – “software and workstations for antiquated technology” hubs that have expertise to share regarding the handling of digital media.<sup>12</sup> The benefit of this plan is that it does not necessitate all repositories becoming technical experts on all varieties of born-digital materials from all eras. This sort of collaboration holds great possibility.

The study presented in this paper used both a survey and semi-structured interviews with archivists to investigate whether and how manuscript repositories are currently managing born-digital materials. While not focusing too closely on the technical issues of born-digital records and not summarizing all the debate in archival literature about issues that relate to processing these materials, the intent of this study has been to pinpoint some of the problems that plague manuscript repositories in particular, to identify some practical steps that should be replicated, and to suggest courses for further study and action.

## **Literature Review**

For its Digital Preservation Outreach and Education program, the Library of Congress defines six core principles for its curriculum.<sup>13</sup> These topics were adapted and expanded by Cal Lee for the Closing the Digital Curation Gap project,

a partnership between the University of North Carolina at Chapel Hill and Jisc (formerly the Joint Information Systems Committee).<sup>14</sup> They provide the framework for a review of scholarship on born-digital materials.

### *1. Prepare*

In her list of steps and strategies to begin addressing the problem of electronic records in manuscript repositories, Elizabeth Dow counsels that a repository should work to develop or amend policies before committing to the long-term preservation of born-digital records.<sup>15</sup> In her 2006 article about the acquisition of the Michael Joyce Papers at the Harry Ransom Center, Catherine Stollar Peters echoes this opinion, seeing policies as evidence of an institutional commitment to the project.<sup>16</sup>

While the theory of developing policies before the acquisition and processing of born-digital materials holds merit, in practice it does not usually seem to be the case. In her 2007 survey of collecting repositories, including both public and private academic institutions and historical societies, Susan Davis found that only 24 percent of the institutions had a policy in place regarding the acquisition of digital records. Of that subset, 57 percent of those policies mirrored the policies for traditional collections. She quotes a respondent from a public university who summarized the situation: “We are passively accepting born-digital materials. We don’t even have a plan for preservation of the digital surrogates we are creating. We barely have enough staff to cover reference and manage limited processing. All planning, policy, etc. take a back seat to day-to-day efforts to keep up with basic activities.”<sup>17</sup> In a similar vein, Ben Goldman, writing about his work at the American Heritage

Center, contends that beginning the work first can be a very valuable means of shaping the necessary policies and procedures for a repository.<sup>18</sup> Some questions cannot be answered (or even anticipated) unless the repository is already doing work with born-digital materials.

## *2. Identify*

The step of identifying includes both determining what born-digital materials might already be within collections and deciding what born-digital materials should be accessioned. In many cases, electronic media have been accessioned in hybrid collections without adequate documentation of their existence. In his case study on the Beinecke Rare Book and Manuscript Library, Michael Forstrom defines such “fugitive media” in this way: “there has been no significant precustodial intervention, the digital content has not been appraised prior to acquisition, and the media is part of a collection consisting chiefly of paper-based materials.”<sup>19</sup> In defining the process of surveying digital materials, Elizabeth Dow acknowledges a physical survey of an accumulated collection is prohibitive; instead, “surveying digital materials depends on determining the context of the materials’ creation and use.”<sup>20</sup>

The broader issue underscored here is determining when born-digital materials should be identified for long-term preservation in a manuscript repository. Elizabeth Dow points out that archives have typically embraced the life cycle model when dealing with paper records, usually acquiring materials only when they have become inactive records. However, due to the “fragility or impermanence of digital documents,” Dow argues the continuum model provides a much more viable representation, which also

“implies that archivists should identify digital materials of archival value and assert some authority over them at creation, or before.”<sup>21</sup> Adrian Cunningham concurs and adds that archivists frequently interact with highly sought after donors before any agreements are signed, so conversing about software platforms and file naming conventions should not be seen as an extraordinary measure.<sup>22</sup> He also argues that “much of the impetus for continuum thinking has come from the emergence of electronic records.”<sup>23</sup>

### 3. *Select*

Given the complexity of the long-term preservation of electronic records, there is an ongoing debate within the archival community about which records need to be maintained in a digital format. In his 2007 book *Records Management*, David Stephens argues that in a business environment, data should only be preserved in a digital format if conversion to an analog format “would severely diminish its value or render it unusable in order to satisfy required (rather than ‘nice to have’) business requirements.”<sup>24</sup> Elizabeth Dow uses colorful imagery to describe the appraisal of electronic records, referring to “the specter of the certain death of digital documents” and suggesting that this threat of the imminent demise of electronic records actually gives a curator more latitude to reject donations of questionable materials due to the effort required to maintain them – though she also cautions that, unlike paper records, collections that are rejected today are not likely to be available for reconsideration in the future.<sup>25</sup>

This leads into another debate raging in the archival community – whether archivists should have precustodial intervention with donors in order to identify records that

should be kept and help ensure that these born-digital materials will persist until they can be deposited in a manuscript repository. Dow acknowledges the side of the debate that worries such interventions might preclude the otherwise unselfconscious documentation by a donor, but she concludes that this risk is better than the alternative of having no viable records to ingest at the end of a person's career.<sup>26</sup> Cunningham argues that "self-conscious record keeping" already exists outside of the realm of electronic records, so this should not be a reason to avoid the precustodial interventions that might ensure the continued viability of this evidence. He also suggests that functional appraisal is the answer to discerning at the stage of creation which digital objects will hold historical significance. And where some discount the viability of precustodial intervention by suggesting it is too time consuming, Cunningham argues that it is merely a reallocation of time that otherwise would have been spent later in the process.<sup>27</sup> Tom Hyry and Rachel Onuf directly counter Cunningham's arguments, suggesting that precustodial interventions would skew appraisal decisions toward individuals who gain fame early in their lives and would force these decisions to occur without the perspective that comes with the passage of time.<sup>28</sup> But a manual recently published by the Society of American Archivists (SAA) supports proactive involvement with donors, arguing that "much of the metadata used by archivists to add value to the digital records and manuscripts is best captured before it comes to the archives."<sup>29</sup>

The digital housekeeping practices of the creators of born-digital materials strongly influence the ability of a repository to determine which digital objects warrant preservation. As a result, repositories such as the Beinecke

Rare Book & Manuscript Library have developed suggested guidelines for authors who intend to deposit their work. The minimum steps identified are: (1) save old physical media that contains unique files, (2) back up files, (3) use consistent file naming conventions, and (4) organize files logically. They go on to suggest guarding against obsolescence, ensuring interoperability, adopting standards put forth by national and international organizations, and ensuring backwards compatibility.<sup>30</sup>

However, manuscript repositories most often receive collections from donors who have not had extensive collaboration with the repository during the creation of the records. Susan Davis points to the complications that arise from receiving electronic records in a multitude of formats and without adequate accompanying metadata.<sup>31</sup> The results of her 2007 survey indicate that repositories may alter their acquisition procedures to reflect the particular concerns raised by born-digital materials by conducting more extensive negotiations with the donors of digital objects, asking for additional documentation, limiting the acceptable formats of electronic records, and specifying software and/or hardware requirements.<sup>32</sup>

Some of the loudest voices arguing for earlier interventions by archivists come from the “new paradigm” theorists, who also tend to favor a documentation strategy for appraisal and the continuum model of records. At the same time, they support a system wherein the records creators remain the custodians of the digital objects and archivists serve as consultants.

#### *4. Get*

During the 1990s, two schools of thought emerged about the appropriate locus of custody of electronic records. Luciana Duranti and Terry Eastwood from the University of British Columbia sought to apply traditional archival and diplomatics theory to electronic records and argued that the archive should serve as the holding place for electronic records, just as it has for centuries for paper records.<sup>33</sup> The Pittsburgh Project headed by David Bearman and Richard Cox argued that there needed to be a “new paradigm” in archival thinking to handle electronic records, embracing the continuum model of records and asserting that a noncustodial role was the only realistic one for repositories.<sup>34</sup> As Cunningham points out, the problem with the noncustodial position is that it does not encompass the need for archives for personal papers: “Governments and organisations may exist for indefinite periods of time or have cooperative successor organisations. Private individuals have an unfortunate habit of dying and leaving relatives who refuse to have any truck with the ongoing custody of the deceased’s records and who, in any case, probably could not be entrusted with the responsibility.”<sup>35</sup>

The 2012 ARL survey posed the question, “Which of the following strategies does your library employ when ingesting born-digital records stored on legacy media?” Choices included storing legacy media as is, developing a collection of legacy hardware, outsourcing, building new systems, and participating in a collaborative. While the respondents to this survey were limited to the members of the ARL, the results certainly indicate that electronic records already are being accessioned by repositories, with most current strategies passive ones.<sup>36</sup>

##### *5. Store*



With all of the attendant complications of preserving and providing access to electronic records, some repositories are choosing to create hard copy records and store those as record copies. Cunningham contends that this is an acceptable solution for some institutions, provided that there is no loss of necessary functionality through the conversion to analog and that the process of creating and maintaining the paper records is cheaper than the alternative digital choice.<sup>37</sup> Dow also suggests that converting born-digital materials to an analog format is a reasonable preservation solution for small repositories, so long as the “essential contextual information” is preserved in this printout. However, she also acknowledges that conversion carries an opportunity cost, for the analog surrogates share none of the functionality of their digital counterparts (e.g., being able to manipulate data in a database or follow a hyperlink in a document).<sup>38</sup> Ben Goldman, however, contends this is not a scalable procedure, for “there is not enough paper in the world to print, en masse, all the electronic records we have acquired (and will likely acquire in the future), nor would the solution even be appropriate for more complex types of digital files, such as databases, Web sites, or multimedia.”<sup>39</sup>

An outgrowth of the debate over the life cycle versus the continuum models of records has been concern over whether traditional methods of archival description can adequately describe electronic records. Kathleen Roe argues for adaptation because the model of the physical arrangement of paper records does not translate directly to born-digital materials. She suggests the records creator needs “to identify the records systems within which electronic records function. This focuses attention on how the intellectual relationships

among databases or electronic files supported an organization's functions and activities."<sup>40</sup>

Philip Bantin provides a useful comparison of the description of electronic records within the continuum and life cycle models. Where proponents of the life cycle model argue that traditional archival description provides the best means of protecting authentic records, advocates of the continuum model suggest four reasons alternative description methods are warranted. For one, effective description should take place during the life of the record, not when it becomes inactive. They also point out that prose descriptions do not reflect the complex relationships of digital objects. They acknowledge that the physical review of files to determine content and context is not viable for handling the scale of records produced in electronic environments. Lastly, they suggest that record system metadata is an existing alternative for description.<sup>41</sup>

This question of whether metadata provides adequate description has received quite a bit of attention in the archival literature.<sup>42</sup> Writing in *Archivaria* in 1993, Margaret Hedstrom explains that description for all types of records should allow users to identify and locate records, understand the record and interpret its content, and establish the authenticity of the record; apart from interaction with users, the description should also help manage the record. She contends that archivists would be better served by capturing metadata generated in the records systems rather than generating it themselves.<sup>43</sup> David Wallace contends that in order for this to happen, archivists need to be involved in the creation of "electronic record-keeping systems" – which he differentiates from "data management" – that prioritize timeliness and reusability of data rather than documenting

transactional evidence.<sup>44</sup> But where Wallace sees metadata capable of doubling as archival description, Heather MacNeil disagrees. She compares metadata to a diary and description to “a biography, that, in narrational style, examines a life already lived, from a perspective broader than that in which it was lived.” She goes on to suggest that the volume of data generated by metadata systems is so vast that it “may in fact obscure, rather than illuminate, the broader administrative context and thereby bias the users’ understanding of the records’ meaning.” MacNeil directly challenges Hedstrom’s argument, contending that using metadata as archival description actually perverts the primary purpose of metadata and thereby “contravenes the archivist’s primary duty to protect and preserve the inherent characteristics of archives – their impartiality, authenticity, and interrelatedness – which derive from the circumstances of their creation.”<sup>45</sup> In analyzing this debate, which took place on the pages of *Archivaria* in 1995, Wendy Duff cautions, “before archivists abandon archival description, they require research that compares the retrieval performance of the two types of systems: one containing descriptions consisting of metadata and the other with descriptions supplied by archivists.”<sup>46</sup> Unfortunately, her directive to focus on users has not been heeded.

## 6. *Protect*

Charles Dollar offers a simple definition of the preservation of electronic records: “ensuring their readability and intelligibility in order to facilitate data exchange over time.” Many of the recommendations he made in 1992 for dealing with technology obsolescence are still embraced today, such as advocating for open systems standards and

identifying migration paths.<sup>47</sup> Dow also identifies five issues that have to be addressed in order to preserve electronic records for the long-term: preserving the hardware, the software, the storage medium, the skills (i.e., being able to use older programs and make sense of stored data), and the information. She goes on to suggest that since it is impossible to anticipate future uses of electronic records by researchers, archivists' goal should be to guarantee the reusability of these records. Accomplishing this goal will necessitate protecting the records from change, ensuring that migrations render documents that are "coherent, reconstructible, and functional," and documenting actions taken both while records reside in a repository as well as at the time of deposit.<sup>48</sup> On a more technical level, David Stephens identifies five types of data preservation practices. They are (1) updating the media on which electronic records are stored, (2) migrating data to new formats, (3) standardizing file formats, (4) recopying media at specified intervals, and (5) emulating the environment in which the digital object was created.<sup>49</sup>

The 2010 report of the Digital Lives project makes several suggestions about future roles for manuscript repositories, one of which pertains to the issue of authenticity. It suggests that repositories may not always have a custodial relationship to digital objects created by individuals and should look to become "guardians of the authenticity of the originals including digital objects in the wild."<sup>50</sup> This parallels the noncustodial model of recordkeeping advocated much earlier by the Pittsburgh Project.

## *7. Manage*

While in the beginning many born-digital collections were treated as special projects, that is an increasingly ineffective strategy. As indicated by the 2012 ARL survey, the “trickle” of electronic records has become a “flood,” so archivists “must develop policies and procedures to operationalize the management of born-digital materials, or we risk losing the record of the recent past.” The respondents to this survey indicated four critical developments that will push the management of born-digital materials from the project phase to the program phase: (1) “Collaborative solutions for dealing with hardware and software obsolescence”; (2) “More, and more appropriate, storage for born-digital materials”; (3) “Automation of as much of the workflow as possible”; and (4) “Asset-level access control to enable tiered access to restricted records.”<sup>51</sup>

There is no doubt that money is a dominant factor in how manuscript repositories choose to handle born-digital materials. As Cal Lee concludes from lessons learned working with electronic records in state government, “resources are limited, meaning is expensive.”<sup>52</sup> The costs include purchasing and maintaining updated technology necessary to preserve and provide access to electronic records, along with hiring staff competent to work with digital objects. Repositories must then find new funding to pay for these costs or alternatively reallocate existing resources in order to provide additional services. Terry Cook identifies “the stark bottom line: unless you can get substantial new financial and human resources, you will need to stop doing something important that you are now doing and reallocate significant resources to electronic records, period. There is no other way.”<sup>53</sup>

David Bearman and Margaret Hedstrom borrow an image from the book *Reinventing Government* to illustrate how they think electronic records can reinvent the archival profession. They conclude, “electronic records can be a vehicle for archives to move from rowing to steering, towards more enterprising and customer driven approaches to service delivery, and towards empowering others to take action in a decentralized records management environment.”<sup>54</sup> Rick Barry also believes that born-digital materials bring the possibility of change, though not necessarily with quite as much control. He asserts that managing born-digital records involves different requirements than traditional paper records: “new skill sets, sophisticated, trustworthy, software tools and a great deal of our only inelastic resource – time – to carry out concentrated planning, stakeholder management, and training efforts, all with ever diminishing levels of human and capital resources being allocated to meet these challenges.”<sup>55</sup>

#### 8. *Provide*

Duff offers an analysis of why providing access has generally been complicated for archives. Speaking at the 2002 DLM-Forum, she suggests that archivists usually focus on the act of record creation rather than on the secondary uses of these records.<sup>56</sup> And, of course, without a consideration of secondary uses, there are no users of manuscript repositories, for the record creators do not typically make frequent use of inactive records they have deposited in a manuscript repository. Writing in 1994, Cunningham asserts that the preservation of electronic records is pointless without adequate provision for user access (along with the requisite training to make good use of

this access). He also suggests that providing “networked access” to remote patrons should soon be a viable option.<sup>57</sup> One repository that is currently offering online access to some born-digital collections is the University of Illinois at Urbana-Champaign. However, in the case of the papers of Stanley Smith, a former chemistry professor, the repository warns the online user that links to some digital documents may no longer work.<sup>58</sup> The results of the 2012 ARL survey indicate that “access to collections is not as fully developed as the management of born-digital content.” The results of this survey go on to suggest that the two biggest access challenges are the sensitivity of materials and the lack of IT infrastructure. Along with this is the concern that automated systems are not capable of dealing with complex access restrictions in the same capacity as traditional reference desk staff.<sup>59</sup> Erway’s 2010 essay raises a related question: “should digital access be subject to the same constraints as analog access?”<sup>60</sup> The documentation from the AIMS project divides access to electronic records into four levels: discover, view, render, and download.<sup>61</sup> Despite numerous recognitions that patron access is the end goal, this element of the workflow seems to be the most difficult to solve.

Wendy Duff has done a lot of writing over the years about archives patrons. Although no one seems to have answered her 1995 call to investigate the viability of metadata as a substitute for archival description, she herself paired with Catherine Johnson in 2002 to write about a subset of archive users, historians. In their analysis of participant comments about how they orient themselves to an archives, Duff and Johnson include a revealing quotation about the value of personal contact with a knowledgeable archivist: “all of the...best digitized sources in the world are

never going to replace that for me.”” The rise of the MPLP (More Product, Less Process) movement perhaps indicates that the days of archivists well-versed in the intricacies of their collections have already passed; nevertheless, Duff and Johnson’s conclusion is still worth acknowledging: “archivists were easier to use than finding aids and could make connections to relevant material in a way that was impossible to replicate in either the printed or online aids.” The relatively unposed and certainly unanswered question is whether researchers will be comfortable transitioning to a relatively unmediated presentation of born-digital materials, which seems to be the model currently gaining traction. While this mode is likely effectual when searching for known materials, the mechanisms for perfecting the recall and precision results of more exploratory searches have not been developed. The conveyance of contextual information should also be addressed; participants interviewed by Duff and Johnson speak both of the need to understand a document in the context of the entire collection, as well as to gain insight into what a particular collection holds and what it lacks and why.<sup>62</sup>

There are three primary methods of providing access to digital objects over time: generating an analog version (i.e., printing a hard copy), migrating the digital object to a format compatible with current computer systems, and emulating the original platform in which the digital object was created.<sup>63</sup> The latter mode in particular seems to be out of the reach of most manuscript repositories today, despite the Digital Lives report referring to emulation as “an essential approach” and “the preferred access route for many eMSS scholars.”<sup>64</sup> Nonetheless, emulation has not taken hold as a preservation strategy. The 2012 ARL survey posed the



question, “Which of the following delivery methods does your library use to provide access to born-digital materials?”<sup>65</sup> The results indicate concretely that only one participating repository practices emulation (see Table 1).

**Table 1. Results of ARL Survey on Methods for Access to Digital Materials**

Online access to a digital repository system	42	66%
In-library access on dedicated computer workstation	31	48%
In-library access using portable media accessed through the users' personal computer	22	34%
Third-party access & delivery system	18	28%
Online access to a file space	15	23%
In-library access to records in an emulated environment	1	2%
Online access to records in an emulated environment	1	2%
We do not provide access at this time	13	20%
Other delivery method	10	16%

Much of the excitement about born-digital records relates to the searchability that characterizes a digital environment. As Dow suggests, the level of search available in electronic records provides “a quality of intellectual access almost impossible to deliver in an analog document.”<sup>66</sup> In response to Davis’ 2007 survey, one public university archivist commented, “I am inclined to accept some digital materials that I might be reluctant to accept in paper format.

This is because ephemeral materials take on new value when they are part of a body of material that can be searched using full-text search engines.”<sup>67</sup> NARA seems to be embracing the greater access that can be provided for electronic records. In a 2009 workshop, Kenneth Thibodeau reports, “NARA has decided that the public will need to go to only one place in ERA [Electronic Records Archives] for access to all records which are publicly available, even when there are some restrictions on content. In the public access part of ERA, anyone will be able to find information about any records we preserve, both traditional and digital, federal, presidential, and those Congressional records we are allowed to release to the public.”<sup>68</sup>

Research about the users of these electronic records is one area in which literature on born-digital materials is lacking. This can somewhat be explained by the relative scarcity of access to these records, along with researchers’ comparatively minimal interest in the time period generally covered by these materials – the last three decades. The white paper produced by the AIMS project acknowledges that the ability to make born-digital materials discoverable and accessible online opens up many possibilities, but doing so also de-personalizes the archival research process by potentially removing the archivist from that process, thereby eliminating one means of ensuring appropriate access and use of materials and increasing “the risk of misuse or abuse of copyrighted or sensitive information.”<sup>69</sup> Despite uncharted terrain, it is vital to consider how born-digital materials held in archives might be used. Eric Ketelaar refers to the “affordances of digital technologies” that “stimulate people to create content differently and to use documents differently in different collaborative and distributed networks.”<sup>70</sup>

Without consideration of the ways that the work of records creators continues to change as well as the ways that the work of records users continues to change, archives themselves could truly become relics of the past.

### **Research Design**

A fairly exhaustive review of the literature about electronic records reveals that there are many voices not currently represented, particularly from the manuscripts community. One way the Manuscript Repositories Section of SAA is attempting to get more involved in electronic records management efforts is through the Jump In Initiative, which challenged SAA members to begin managing born-digital content and specified steps drawn from Erway's *You've Got to Walk Before You Can Run* report.<sup>71</sup> In addition to conducting a broad literature review that included case studies from manuscript repositories, the author chose to survey the membership of the SAA Manuscript Repositories Section discussion list for this study. This survey began as a request on Manuscripts Repositories Section listserv that repositories not currently processing born-digital materials answer a two-question survey online and that repositories already processing electronic records contact me to set up a time for an interview.<sup>72</sup>

Five members took the survey for repositories that are not currently processing born-digital materials. The intent of this survey was to identify potential roadblocks to handling these materials.<sup>73</sup> Five members who are currently processing electronic records provided feedback through email or phone interviews; interviews with four additional archivists were arranged through other contacts. The intent of these interviews was to ferret out policies or procedures

that are working effectively, as well as to ascertain challenges that persist.<sup>74</sup> There is arguably a self-selection bias to the pool of respondents, as several did mention having a connection to the School of Information and Library Science at the University of North Carolina at Chapel Hill or to study advisor Jackie Dean. An additional email was sent to those who registered for the Jump In Initiative in the hope of gaining an early glimpse of their findings and to determine their motivations for participating in the initiative.<sup>75</sup> Out of the thirty-three people registered, seven provided additional feedback about their work on this project.

## **Findings**

### *1. Survey of Repositories Not Currently Processing Born-Digital Material*

Three of the responses to the survey came from special collections repositories at a university; one came from a government institution; and one came from a public library. The first question of the survey asked respondents to identify factors that have limited the ability of their manuscript repository to process born-digital records. The options provided were training, costs, concerns about providing access, time, and inadequate administrative support. Each of these factors was rated by the respondents as having some significance in their inability to begin processing electronic records. Each possible response was weighed from 1 to 4, from no significance to highest significance. Given this framework, inadequate administrative support returned the highest result with an average score of 3.4. The next highest result was a 3.2

average score for training to know how to handle born-digital records.

**Table 2. Question 1 Responses**

#	Question 1	Small	Significant	Monumental	Total	Mean
1	Training to know how to handle born-digital records	0	4	1	5	3.20
2	Cost of handling born-digital records	1	4	0	5	2.80
3	Concerns about providing patrons access to born-digital records	2	3	0	5	2.60
4	Time required to process born-digital records alongside backlog of paper records	1	4	0	5	2.80
5	Inadequate administrative support	1	1	3	5	3.40

The second question asked the respondents to consider what scenarios would facilitate the processing of born-digital records by their manuscript repository. Once again, the responses were values on a Likert scale from 1 to 4, from no significance to highest significance. By far the option receiving the highest score (3.8) was the development of acquisition, preservation, and access policies for born-digital materials. And by far the option receiving the lowest

**Table 3. Question 2 Responses**

#	Question 2	Not an aid	A small aid	A significant aid	An extremely important solution	Total Responses	Mean
1	Being able to bring in a consultant to provide assistance in establishing a workflow for processing born-digital records	0	1	4	0	5	2.80
2	Being able to study established workflows from other similarly sized repositories to learn from their processes	0	1	3	1	5	3.00
3	Being able to acquire grant funding to help cover costs	1	0	1	3	5	3.20
4	Being able to provide training for reference staff in how to facilitate patron access to born-digital records	0	1	4	0	5	2.80
5	Being able to provide training for patrons in how to access born-digital records	0	5	0	0	5	2.00
6	Being able to hire an electronic records archivist who could focus only on the processing and access of born-digital records	1	0	2	2	5	3.00
7	Being able to develop an institutional policy governing the acquisition, preservation, and access of born-digital materials	0	0	1	4	5	3.80

score (2.0) was the provision of patron training in how to access these records.

## *2. Interviews with Repositories Currently Processing Born-Digital Materials*

Seven archivists shared their time and expertise through interviews.<sup>76</sup> Several of these repositories, specifically the University of Mississippi Archives and Special Collections and the American Heritage Center at the University of Wyoming, began working with electronic records because they anticipated acquiring collections with significant amounts of digital content, so they wanted to be proactive and have workflows in place to handle those born-digital materials. At the New York Public Library (NYPL), work in this arena was sparked with the hire of a digital archivist, as well as the anticipation of receiving more born-digital collections. Much of the work at the American Heritage Center, such as designing a workflow and completing an inventory, also began after the hiring of a digital programs manager.

Glynn Edwards made a presentation at the 2012 Rare Books and Manuscripts Conference where she provided an overview of the born-digital workflow in Special Collections at the Stanford University Libraries. It includes using forensic software such as FTK Imager to create checksums for authenticity and to create directory listings along with Archivists' Toolkit for registering the objects and creating finding aids.<sup>77</sup> In the case of the Stephen Gould Papers, Stanford chose to assert intellectual control over the electronic records by mirroring their organization to that of the physical files.<sup>78</sup> Don Mennerich at the New York Public

Library is also using forensic tools to extract metadata and assert intellectual control over digital objects.

Some hesitated to compare the processing time of born-digital collections to analog collections because there is no precise metric for doing so, but Jeff Thomas of The Ohio State University Libraries contends that “processing digital documents consumes more time than paper records. Computer files simply take a lot longer to browse through than flipping through paper.”<sup>79</sup> While several repositories began their work by generating item-level metadata for digital objects, they have come to the conclusion that this is not a scalable approach; Thomas is especially adamant that the processing and arranging of electronic records must occur no lower than the folder level.

For most of these repositories, providing access is the last piece of the puzzle. To this point, there has been no notable pressure from patrons to provide access to born-digital materials online, likely because these collections tend to be under-described and therefore users have to locate these resources in finding aids somewhat serendipitously. Two of the repositories surveyed are creating PDF access copies of documents. In the case of the Ohio Congressional Archives, this decision was made both because the PDF format is more secure than the Microsoft Word format in which documents were received and because the PDF files can then be grouped into portfolios for easy online access and keyword searching. The digital objects in the Stanford collections are searchable on a workstation in the reading room. This computer is not on a network and there are no ports for external drives, so patrons must flag any items that they wish to print and get the assistance of the staff. The NYPL uses QuickView Pro for file viewing and migrates



Microsoft Word documents to ensure they do not lose their search functionality; it is in the process of setting up a media workstation in the reading room. The library may in the future try to virtualize this workstation for remote access.<sup>80</sup> While the NYPL does make an effort to remove personally identifying information such as that in medical records using Bulk Extractor to redact information out of the disk image, they also recognize that it is impossible to redact everything. In the case of digital photographs in the special collections of the University of North Carolina at Chapel Hill (UNC-CH), the workflow is to upload records into the CONTENTdm system for access but to suppress the image when the donor agreement with a photographer requires permission for a patron to use an image. Patrick Cullom of UNC-CH Special Collections acknowledges that young researchers in particular presume that they should have a right to access digital “stuff” merely because the technology makes it easily available, but he goes on to point out that archives have a responsibility, as they always have, to protect the items deposited with them; therefore, an ability to access a digital object does not necessarily equate to a right to access that object.

In completing its work on the AIMS project, Stanford Libraries wrote guidelines for creating donor agreements; these guidelines point out the importance of documenting issues relating to ownership, exclusivity, and preservation, to name a few.<sup>81</sup> With all of the interview subjects, policy work has followed the creation of a basic workflow process. Mark Greene has long advocated early involvement with donors, and he sees no reason not to transfer that practice to donors of born-digital materials. Writing about his time at the American Heritage Center, Ben

Goldman speaks of the importance of conducting preacquisition appraisals. Jeff Thomas strongly favors precustodial interventions with Congressional offices to educate them about the importance of creating an organized foldering system and following filenaming conventions.<sup>82</sup>

One of the suggestions for improving the management of born-digital materials is to create an institutional repository that could be responsible for the long-term maintenance of ingested digital objects. Another very practical suggestion is to accumulate now the equipment that will be necessary to access files later; for instance, UNC-CH is building a “Frankenstein” machine that will have the capability to access files from various types of digital media cards and other media formats. Another surveyed repository recognizes that more money and more staff are necessary to handle born-digital materials effectively. One respondent points out the need to get electronic records documented in the processing manual for his repository. And Mark Greene is emphatic that more people need to be competent and comfortable in working with born-digital materials rather than isolating that expertise, but it seems more common that fewer people are involved in work with electronic records during the initial planning and implementing stages.

Given that for many years NARA provided leadership for the archival profession, archivists at the two most recent presidential libraries were contacted to find out how they handle born-digital materials and to determine if there are any lessons that can be applied to other types of repositories.<sup>83</sup> While the Clinton Presidential Library does have a database of emails (plus attachments) generated by the Clinton White House from 1993 to 2001, Adam Bergfeld explained they cannot provide electronic access to these

materials for security reasons. Materials are accessed through Freedom of Information Act requests, at which point he searches the repository for relevant records (both electronic and analog) and provides paper copies to the researcher.<sup>84</sup> However, Sarah Ticer at the George W. Bush Presidential Library explained their goal is to make processed born-digital records available through NARA's Online Public Access catalog. So even though there will not be direct links from the online finding aids hosted on the Bush Library web site, there will be a mechanism for online delivery of records.

### *3. Feedback from Jump In Initiative Participants*

Five of the manuscript repositories that provided feedback are housed in universities. Another is a historical society, and one is a religious organization.<sup>85</sup> There were numerous explanations of their motivations to participate in the Jump In Initiative. Respondents A and G both mentioned the importance of knowing participants are a community of other archivists working through the same issues at the same time who can be looked to for guidance and support. Respondent A elaborated to say, "now that I've said I'll do this and I am part of this group, I feel obligated to finish." Respondent B asserted that having SAA sponsor this activity gave an aura of "credibility/authenticity/authority" that helped her sell the project to her superiors. Respondent C suggested the requirements of the initiative were simple enough that "there is little to lose and easily something to gain." Respondents B and D both acknowledged the possibility of winning tuition to a Digital Archives Specialist (DAS) class helped seal their decisions to participate. Respondents A and E mentioned having a deadline can be helpful when confronting a difficult task. Respondents B, E,

and F all indicated they recognized they needed to conduct a survey of the electronic media in their collections, and this initiative gave them the incentive.

The early results of the surveys of computer media types were also wide-ranging. Respondent A actually found fewer computer media than presumed. Respondent B, on the other hand, found many more storage media formats than expected, but was relieved to find that more of them are of the CD and DVD variety rather than the more difficult to access 3.5-inch floppy disk. Several respondents found “fugitive” media in collections that had not been properly identified in finding aids or other accessioning materials. Respondent C admitted “I’m afraid that we have in the past adopted the ‘file it and forget it’ approach to the problem; we’re in for nothing but surprises in earlier acquisitions.” Along with media that were not counted in the finding aids, Respondent F found instances where born-digital materials were printed at the time of donation, and the physical media were never deposited. Respondent G indicated that simple searches of the finding aids for terms like “CD” or “computer” were not sufficient to find all of the computer media in their collections. Respondent D found a plethora of CDs and DVDs, many of which are “commercial appearing disks” likely to raise copyright concerns.

More of the respondents (A, B, C, D, and F) indicated that they tend to actually work with born-digital materials first before designing overarching policies that apply to these records. A few of the institutions already have some relevant policies in place; for instance, a retention schedule governs the accessions of Respondents A and E. Respondent G indicated that the repository is “looking more at the big picture and working our way down,” and therefore

has incorporated language about born-digital items into donor agreements and developed a digital strategy that outlines mechanisms for establishing a “trustworthy” repository. Respondent F mentioned experimentation with Duke Data Accessioner and work to construct a “forensic and quarantine machine to use as a point of ingest.” Finally, although policy formation is not necessarily the first piece of the puzzle for Respondent E, work for the Jump In Initiative is anticipated to be helpful in developing a protocol for accessions. For instance, legacy electronic records for which no appraisal decisions were truly made will in the future have a framework to help determine which digital objects are truly worthy accessions, thereby limiting the electronic records on which she needs to perform preservation measures. Several repositories deal with born-digital materials on an as-requested basis; for Respondent D this means if there are no requests, there likely will be no preservation steps taken, and if there are requests for immediate use, this may entail serving content from the original disks. Respondent C indicated that the repository has a “standing practice of having surrogates made only when readers request access to material on obsolete media,” although they do “proactively create surrogates” for some “very high-use collections.”

## **Discussion**

Anne Gilliland-Swetland best describes the feelings of many archivists when confronted with born-digital materials: “confused, anachronistic, insecure, even stupid. Like a rabbit out of its burrow on a dark night, many an archivist, faced with venturing into the realm of electronic records, has found herself or himself frozen in the lights of

oncoming traffic, unable to move either forward or backward, doomed to be roadkill on the information superhighway.”<sup>86</sup> Ben Goldman describes the quest to resolve the issues surrounding born-digital materials as a “Quixotic one,” with archivists waiting “for that one perfect, affordable, all-encompassing solution for electronic records.”<sup>87</sup> In his musings on the value and values of archivists, Mark Greene interjects a thought that has interesting application to born-digital materials. He suggests that archivists “tend to focus too much on our processes and not enough on our purpose.”<sup>88</sup> An earlier article by Greene provides a broad answer to the question of what purpose archives serve: “the archival mission is about meaning.”<sup>89</sup>

The time for passivity has elapsed. Even writing in 1994, Adrian Cunningham used the metaphor of a ticking time bomb with regards to electronic personal records in a precustodial environment, but he concludes that “the approach has been to ignore it in the hope that by the time the suspect device is offered for transfer someone will have discovered an easier way of defusing it than is currently available. This approach may be tantamount to the reckless endangerment of both the records themselves and to the very future of those institutions that collect personal records.”<sup>90</sup> Patrick Cullom adds an anecdote from the visual materials realm, suggesting that the archives profession tends to be wary of moving too fast with change because they have been burned in the past with decisions, such as switching from nitrate to safety film.<sup>91</sup> But the luxury of a wait-and-see attitude has long since passed.

In his 2009 article, Adrian Cunningham concludes with a simple to-do list for the archival profession: “conduct more research into the dynamics of personal record keeping,

the societal warrants for personal record keeping, and the functional requirements for evidence in personal record keeping.”<sup>92</sup> Yet these directives are probably still a little too heavily tilted toward theory than practice. This survey returned sharp differences between those repositories engaged in curating born-digital materials and those repositories yet to enter the realm of born-digital archivy; for example, the repositories not currently working with born-digital materials indicated that they think policies should be in place before processing records, while those repositories already engaged in the work find it more effective to divine appropriate policies only after understanding the various attendant issues of handling born-digital records. Given these differences, it seems imperative that the archival literature begins to reflect more of the common sense approaches developed by those in the trenches. Just as importantly, there must be more research into the users of born-digital materials. As Ian Anderson concludes, “if archives are to maintain their high standards of service in the digital age, it is fundamental that these are based on a thorough understanding of users’ information-seeking behaviour and requirements.”<sup>93</sup>

An invitational symposium at the University of Maryland in May 2010 entitled “Computer Forensics and Cultural Heritage” prompted the generation of a list of recommended next steps that included policy frameworks, collaboration, new tools, training, and case studies.<sup>94</sup> Four years later, some work has been accomplished on these steps, but there is still much to do. One example of a project that is attempting to develop new tools is the BitCurator Project. The BitCurator Project aims to help libraries, archives, and museums (LAMs) in “(1) integrating digital forensics tools

and methods into the workflows and collection management environments of LAMs and (2) supporting properly mediated public access to forensically acquired data.<sup>95</sup> While SAA is maintaining an online portal of case studies concerning born-digital materials held in campus archives, most of these do not directly relate to the types of records commonly collected by manuscript repositories.<sup>96</sup> Noticeably absent from the focus of each of these efforts is attention on the users of electronic records.

Writing in 1998, Philip Bantin identifies the “new skills” that will help archivists with electronic records, including “a basic knowledge of how automated systems are created and work.”<sup>97</sup> While there has been much written on these topics in the intervening years and there have been some efforts to address the educational needs of archivists through programs like SAA’s Digital Archives Specialist (DAS) certificate or the DigCCurr program at UNC-CH, these skills remain outside of the grasp of most current archivists in manuscript repositories.

The scientific community has already begun addressing many of the issues surrounding the preservation of born-digital materials. For instance, many grant funders now require data sets to be made public.<sup>98</sup> Unfortunately, according to the 2008 UK Research Data Service feasibility study, the arts and humanities field tends to re-use research data in a manner that differs from the sciences, so the models established by repositories of scientific data may not directly translate to manuscript repositories.<sup>99</sup> Nonetheless, the principle of engendering cooperation among records creators, publishers, other organizations, and data repositories bodes well for the long-term preservation of digital objects. For instance, perhaps archivists could initiate



an alliance with the writers' guild and discuss what sorts of drafts and correspondence should be preserved.

Richard Pearce-Moses used his presidential address to the 2006 annual meeting of the Society of American Archivists to inspire the masses to confront the digital era, concluding with an image of archivists as pioneers on the digital frontier, taking risks in order to preserve our documentary heritage.<sup>100</sup> Given this inspiring image, the question that remains is whether the archival community agrees on this vision of where it needs to go. If the archival profession is dedicated to the long-term preservation of born-digital materials, the literature convincingly identifies the practical issues that need to be addressed in a coherent, unified manner:

- determine a best practice for acquiring born-digital materials (e.g., by transfer of physical media or by disk image created by repository staff)
- determine a method for protecting digital objects, including documentation that can be used for authenticity
- determine a method for appraising and acquiring electronic records, including whether these should occur at regular intervals or once the records become inactive
- determine a mechanism for interacting with potential donors of born-digital materials, including written guidelines of preferred formats and suggested file naming conventions
- determine how to handle the interpretation and application of copyright protection to born-digital objects

- write new acquisition and appraisal policies and donor agreements that incorporate issues unique to born-digital materials

Many resources already exist that can help resolve these issues – it is merely a matter of summoning the collective will to make the decisions that will ultimately benefit all constituencies of manuscript repositories. For example, the 2011 “Managing and Sharing Data” report by the UK Data Archive provides a useful one-page data management checklist.<sup>101</sup> The Digital Curation Centre has been collecting and creating resources for a decade, and their web site includes briefing papers, how-to guides, and a data management planning online tool.<sup>102</sup> The Consultative Committee for Space Data Systems has produced extensive specifications for the Open Archival Information System, and with its acceptance as ISO 14721:2012, this document provides a common framework and terminology for archives that are providing for the long-term preservation of digital objects.<sup>103</sup> The Section 108 study group that was convened by the National Digital Information Infrastructure and Preservation (NDIIP) program of the Library of Congress and by the U.S. Copyright Office has provided a useful commentary on the application of copyright to digital objects.<sup>104</sup> OCLC and the Center for Research Libraries developed criteria and a checklist for measuring trustworthy repositories.<sup>105</sup>

Although it can be tedious, time-consuming work, the policy piece of this problem is actually the easiest to remedy. Just as with the adoption of worldwide description standards, there is much to be gained from an approach that can be embraced by all sizes and types of repositories. Two

issues raised in this study that do not yet have clear solutions are how to provide access to born-digital materials and how to engender administrative support for this work. Perhaps the key to resolving these issues is to recognize that they are intertwined. Cal Lee argues one approach to preserving the layers of meaning held in digital materials is to make the information they possess useful, but this utility is not easily measured when born-digital materials are either undiscoverable or inaccessible.<sup>106</sup> In a time when both public and private funding sources are increasingly limited, it is imperative to demonstrate the positive impact a resource can have on vital constituents in order for that resource – a manuscript repository – to be guaranteed the ongoing administrative support necessary for its long-term health.

## **Conclusion**

Manuscript repositories occupy a unique position in the spectrum of libraries and archives. Unlike repositories where holdings may be duplicates of those found in other institutions or may be of a physical nature that requires little ongoing care, the holdings of manuscript repositories lie clearly in the high stewardship, high scarcity quadrant. The results of the 2012 ARL survey also found that “few of the solutions developed to date have been transferable between institutions.”<sup>107</sup> However, the problems of handling born-digital materials are not unique, and the profession would be well served by finding a space for collaboration to solve these thorny issues. In the words of Don Mennerich, digital archivist for Manuscripts and Archives at the NYPL, “local practice is the enemy.”<sup>108</sup> Margaret Hedstrom warns that “archivists should avoid becoming attached to a model or a formula, because the state of the technological evolution and

the nascent response by archivists do not yet permit conclusive answers.”<sup>109</sup> But as already admitted, these sorts of acknowledgments of the fluidity and complexity of handling electronic records have only served to sanction a dereliction of duty when it comes to putting systems in place that can preserve born-digital materials for the future.

Whether explicitly stated or implicit in interviews and case studies, it seems clear that many manuscript repositories are trying to approach born-digital materials in a manner similar to the way they have handled digitized analog materials. For a multitude of reasons, this is not an appropriate approach. As Liz Bishoff points out, “digital preservation is an ongoing process rather than an event-driven process.”<sup>110</sup> More often than not, digitization has occurred as a limited-scope project, often through grant funding for “boutique” projects that highlight a special holding of a repository as a means of calling attention to the collection. Mark Greene suggests that the tendency to approach born-digital collections in the same manner that digitized collections have been handled will lead to paralysis and serves as evidence that the archival profession is a slow learner, for even in digitization there already should have been a move away from the boutique model.<sup>111</sup> There apparently exists an unhealthy competition for resource allocation between digitized materials and born-digital material; in the section of the 2012 ARL survey dedicated to access and discovery challenges, a respondent indicated that “we often focus on digitizing collections and providing access to those before we can work with the born-digital content.”<sup>112</sup> Yet there exists one dramatic difference between the common approach to digitization and the common approach to born-digital materials: where the primary focus

of the former efforts was to provide access to unique materials, even to patrons who might not be able to visit in person, the primary focus of handling born-digital materials remains mired in a basic level quandary about how to preserve the bits. Until the access piece can be determined, manuscript repositories run the risk of devaluing electronic records.

Does the lack of tangibility and other sensory inputs make born-digital records harder to adequately process and preserve – or does it at least generate less of a visual trigger that there is pressing work to be accomplished? Does the sheer quantity of electronic records make them seem less in line with the mission of a manuscript repository and more appropriate for oversight by an institutional repository? Are there ways in which manuscript repositories can work together to solve some of the problems of born-digital materials, whether by setting up SWAT sites or sharing policies and workflows that can form the backbone of best practices? And most importantly, will the focus of the archival community on born-digital materials ever shift from preservation to access? Obviously, there is still much research that needs to be accomplished in this arena. Perhaps manuscript repositories should look outside of their usual realm to the work being done with repositories of scientific data or to the types of uses of archival records being designed by digital humanists. Based on the positive feedback that the Jump In initiative has generated by creating a sense of community, one simple solution would be to perpetuate this feeling of communal responsibility by establishing mentoring partnerships between manuscript repositories more comfortable with handling born-digital records and those less practiced. Even if standards are

developed, the realm of electronic records is one that will constantly be in flux due to changes in technology, so establishing a support system with other archivists who are facing similar challenges could be a valuable means of preventing the obstacles from appearing insurmountable. Most importantly, if the archival community could embrace a vision of responsibility to provide users with both analog and born-digital materials, figuring out the path to take will be much simpler.

## NOTES

1. Jeff Rothenberg, "Ensuring the Longevity of Digital Documents," *Scientific American* 272 (January 1995): 42.
2. Elizabeth H. Dow, *Electronic Records in the Manuscript Repository* (Lanham, MD: Scarecrow Press, 2009), xiv.
3. Jeff Rothenberg, *Ensuring the Longevity of Digital Information* (Washington, DC: Council on Library and Information Resources, 1999), 3, <http://www.clir.org/pubs/archives/ensuring.pdf>.
4. Kenneth Thibodeau, "Preserving Digital Memory at the National Archives and Records Administration of the U.S." (paper presented at the Workshop on Conservation of Digital Memories, Second National Conference on Archives, Bologna, November 20, 2009), 1, <http://www.mybestdocs.com/thibodeau-k-preser-dig-memNARA091120.pdf>.
5. Margaret Hedstrom, "Electronic Records Program Strategies: An Assessment," in *Electronic Records Management Program Strategies*, ed. Margaret Hedstrom (Pittsburgh: Archives & Museum Informatics, 1993), 1, 4.
6. Adrian Cunningham, "The Archival Management of Personal Records in Electronic Form: Some Suggestions," *Archives and Manuscripts* 22 (May 1994): 95.

7. Matthew G. Kirschenbaum, Richard Ovenden, and Gabriela Redwine, *Digital Forensics and Born-Digital Content in Cultural Heritage Collections* (Washington, DC: Council on Library and Information Resources, 2010), vii, <http://clir.org/pubs/reports/pub149/pub149.pdf>.
8. Dow, xii.
9. Laura A. Millar, *Archives: Principles and Practices* (New York: Neal-Schuman Publishers, 2010), 207.
10. Jackie M. Dooley and Katherine Luce, "Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives," *OCLC Research*, October 2010, 13, 60, <http://www.oclc.org/resources/research/publications/library/2010/2010-11.pdf>.
11. Ricky Erway, *You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media* (Dublin, OH: OCLC Research, 2012), 3-4, <http://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf>. Erway elaborates on each of these steps, and step two includes processes such as identifying file formats and calculating the overall size of the digital collection. She also includes on page 5 eleven technical steps to follow with physical media.
12. Ricky Erway, *Swatting the Long Tail of Digital Media: A Call for Collaboration* (Dublin, OH: OCLC Research, 2012), 3, <http://www.oclc.org/content/dam/research/publications/library/2012/2012-08.pdf>.
13. "Digital Preservation & Outreach Curriculum," Library of Congress, <http://www.digitalpreservation.gov/education/curriculum.html>.
14. "Getting Started Guides," Digital Curation Exchange, <http://digitalcurationexchange.org/gp>.
15. Dow, 15.
16. Catherine Stollar Peters, "When Not All Papers Are Paper: A Case Study in Digital Archiving," *Provenance* 24 (2006): 34.
17. Susan E. Davis, "Electronic Records Planning in 'Collecting' Repositories," *American Archivist* 71 (Spring/Summer 2008): 178, 180.

18. Ben Goldman, "Bridging the Gap: Taking Practical Steps Toward Managing Born-Digital Collections in Manuscript Repositories," *RBM: Journal of Rare Books, Manuscripts, and Cultural Heritage* 12, no. 1 (2011): 21.
19. Michael Forstrom, "Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare Book and Manuscript Library," *American Archivist* 72 (Fall/Winter 2009): 461.
20. Dow, 2.
21. Dow, 8.
22. Cunningham, "Archival Management of Personal Records," 101. For reference, Cunningham also provides a very succinct summary of the continuum model in a footnote in his chapter "Ghosts in the Machine: Towards a Principles-Based Approach to Making and Keeping Digital Personal Records," in *I, Digital: Personal Collections in the Digital Era*, ed. Christopher A. Lee (Chicago: Society of American Archivists, 2011), 89.
23. Adrian Cunningham, "Waiting for the Ghost Train: Strategies for Managing Electronic Personal Records Before It Is Too Late," *Archival Issues* 24, no. 1 (1999): 58.
24. David O. Stephens, *Records Management: Making the Transition from Paper to Electronic* (Overland Park, KS: ARMA International, 2007), 237.
25. Dow, 3.
26. *Ibid.*, 99. She also later admits that donors already shape donations by what they include and what they exclude from collections in any format. *Ibid.*, 117.
27. Cunningham, "Waiting for the Ghost Train," 60-62.
28. Hyry and Onuf, 43.
29. J. Gordon Daines, "Processing Digital Records and Manuscripts," in *Archival Arrangement and Description*, ed. Christopher J. Prom and Thomas J. Frusciano (Chicago: Society of American Archivists, 2013), 95.
30. Beinecke Rare Book & Manuscript Library, "Authors' Guidelines for Digital Preservation," Yale University, <http://www.library.yale.edu/~nkuhl/AuthorsGuidelines.pdf>. For an draft version of the Born Digital Archival Acquisition Collection & Accession



- Guidelines from the Beinecke, see Naomi L. Nelson et al., *Managing Born-Digital Special Collections and Archival Materials*. SPEC Kit #329 (Washington, DC: Association of Research Libraries, 2012), 125.
31. Davis, 169.
  32. Ibid., 182. For a good example of a donor agreement addendum for electronic records from the David M. Rubenstein Rare Book and Manuscript Library at Duke University, see Nelson et al., 122-23.
  33. See, for example, Luciana Duranti, "Concepts and Principles for the Management of Electronic Records, or Records Management Theory is Archival Diplomats," *Records Management Journal* 20, no. 1 (2010): 78-95. This article is a re-publication of a 1999 article.
  34. See, for example, David Bearman, "An Indefensible Bastion: Archives as a Repository in the Electronic Age," *Archives and Museum Informatics Technical Report* 13 (1991): 14-24.
  35. Cunningham, "Archival Management of Personal Records," 99. Although Cunningham does an about-face several years later to embrace the distributed custody model for government and business records, he still maintains personal records will most often need to be taken into archival custody. See Adrian Cunningham, "Journey to the End of the Night: Custody and the Dawning of a New Era on the Archival Threshold," *Archives and Manuscripts* 24 (November 1996): 312-21.
  36. Nelson et al., 35.
  37. Cunningham, "Archival Management of Personal Records," 103.
  38. Dow, 62-63.
  39. Goldman, 14. Given the 1993 decision by the D.C. Circuit Court of Appeals, where the court dismissed the argument by the Reagan and Bush administrations that preserving hard copies of emails satisfactorily maintained the record, it is interesting that this line of thinking has persisted. The opinion states, "if only the hard copy is preserved in such situations, essential transmittal information relevant to a fuller understanding of the context and import of an electronic communication will

- simply vanish.” United States Court of Appeals for the District of Columbia Court, Case No. 93-5002. Scott Armstrong et al v. Executive Office of the President. Decided 13 August 1993. [https://w2.eff.org/legal/cases/PROFS\\_case/appeal.decision](https://w2.eff.org/legal/cases/PROFS_case/appeal.decision).
40. Kathleen D. Roe, *Arranging & Describing Archives & Manuscripts* (Chicago: Society of American Archivists, 2005), 64.
  41. Philip C. Bantin, “Strategies for Managing Electronic Records: A New Archival Paradigm? An Affirmation of Our Archival Traditions?” *Archival Issues* 23, no. 1 (1998): 27-29.
  42. Elizabeth Dow identifies the three critical features of metadata – it describes content, context, and structure. Dow, 33.
  43. Margaret Hedstrom, “Descriptive Practices for Electronic Records: Deciding What is Essential and Imagining What is Possible,” *Archivaria* 36 (Fall 1993): 55, 58.
  44. David A. Wallace, “Managing the Present: Metadata as Archival Description,” *Archivaria* 39 (Spring 1995): 18.
  45. Heather MacNeil, “Metadata Strategies and Archival Description: Comparing Apples to Oranges,” *Archivaria* 39 (Spring 1995): 25, 27.
  46. Wendy Duff, “Will Metadata Replace Archival Description: A Commentary,” *Archivaria* 39 (Spring 1995): 37.
  47. Charles M. Dollar, *Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods* (Macerata, Italy: University of Macerata, 1992), 68, 72.
  48. Dow, 22, 31-32.
  49. Stephens, 238-47. Grimard argues that migration should occur at least at ten year intervals. Jacques Grimard, “Managing the Long-term Preservation of Electronic Archives or Preserving the Medium and the Message,” *Archivaria* 59 (Spring 2005): 166. Standardizing file formats is also commonly referred to as normalization, although Laura Millar refers to it as “migration on ingest.” Millar, 217. Stephens’ use of media recopying as a term is less common than the term refreshing. He

- mentions emulation as a future preservation solution, one that he describes as a “Digital Rosetta Stone.”
50. Jeremy Leighton John, Ian Rowlands, Peter Williams, and Katrina Dean, “Digital Lives: Personal Digital Archives for the 21st Century >> an Initial Synthesis” (2010), vii, <http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf>.
  51. Nelson et al., 11, 18.
  52. Christopher A. Lee, “Guerilla Electronic Records Management: Records Management Lessons Learned,” *Records and Information Management Report* 18, no. 5 (May 2002): 7.
  53. Terry Cook, “Byte-ing Off What You Can Chew: Electronic Records Strategies for Small Archival Institutions,” *Archifacts* (April 2004): 5, <http://www.aranz.org.nz/includes/download.aspx?ID=111276>.
  54. David Bearman and Margaret Hedstrom, “Reinventing Archives for Electronic Records: Alternative Service Delivery Options,” in *Electronic Records Management Program Strategies*, ed. Margaret Hedstrom (Pittsburgh: Archives & Museum Informatics, 1993), 98.
  55. Rick Barry, “Opinion Piece – Electronic Records: Now and Then,” *Records Management Journal* 20, no. 2 (2010): 164.
  56. Wendy Duff, “Understanding the Information-Seeking Behaviour of Archival Researchers in a Digital Age: Paths, Processes and Preferences,” in *Proceedings of the DLM-Forum 2002* (Luxembourg: Office for Official Publications of the European Communities, 2002), 334.
  57. Cunningham, “Archival Management of Personal Records,” 103.
  58. Born-digital and digital surrogates from Stanley Smith Papers, University of Illinois Archives, <http://archives.library.illinois.edu/e-records/index.php?dir=University%20Archives/1505050/>. Based on the work at this repository, Chris Prom wrote a blog entry about their preservation and access plan. Chris Prom, “Simple E-Records Preservation and Access Plan,” *Practical E-Records* (blog), 2011, <http://e->

- records.chrisprom.com/recommendations/supported-formats/simple-e-records-preservation-and-access-plan/.
59. Nelson et al., 15, 17.
  60. Ricky Erway, *Defining "Born-Digital"* (OCLC Research, November 2010), 4, <http://www.oclc.org/resources/research/activities/hiddencollections/borndigital.pdf>.
  61. AIMS Work Group, *AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship* (January 2012), 42, [http://www2.lib.virginia.edu/aims/whitepaper/AIMS\\_final.pdf](http://www2.lib.virginia.edu/aims/whitepaper/AIMS_final.pdf). There is also a useful table found on pages 51-55 that describes various access options as well as the factors that should be considered when deciding which options to provide for each access element.
  62. Wendy M. Duff and Catherine A. Johnson, "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives," *Library Quarterly* 72, no. 4 (Oct. 2002): 484-85, 488.
  63. For an example of a hybrid collection where the repository chose to print hard copies of electronic records, see the James Welch Papers. James Welch Papers, Yale Collection of American Literature, Beinecke Rare Book and Manuscript Library, <http://hdl.handle.net/10079/fa/beinecke.welch>. For a collection where the electronic records are segregated in the finding aid, see the George Whitmore Papers. George Whitmore Papers, Yale Collection of American Literature, Beinecke Rare Book and Manuscript Library, <http://hdl.handle.net/10079/fa/beinecke.whitmore>.
  64. John et al., xiii.
  65. Nelson et al., 71.
  66. Dow, 25.
  67. Davis, 185.
  68. Thibodeau, 8. The description of NARA's prototype Online Public Access can be read at <http://www.archives.gov/research/search/about-opa.html>.
  69. AIMS, 56.
  70. Eric Ketelaar, "Archives in the Digital Age: New Uses for an Old Science," *Archives & Social Studies* 1, no. 0 (March 2007): 174.

71. Manuscript Repositories Section, "Jump In Initiative," Society of American Archivists, <http://www2.archivists.org/groups/manuscript-repositories-section/jump-in-initiative>.
72. The full text of the message can be viewed in Appendix A of the complete version of this paper, available at <https://cdr.lib.unc.edu/record?id=uuid%3aeafafc18-bf84-4aaa-aba8-97c8e4fd18d8>.
73. The instrument that was mounted through Qualtrics can be viewed in Appendix B of the complete version of this paper.
74. While some questions were tailored to the particular collection, the general questions that were asked of all archivists experienced in processing born-digital materials can be seen in Appendix C of the complete version of this paper.
75. The text of this message can be seen in Appendix D of the complete version of this paper.
76. Kathryn Michaelis, email message to author, 7 February 2013; Glynn Edwards and Laura Williams, call with author, 8 February 2013; Don Mennerich, call with author, 14 February 2013; Mark Greene, call with author, 18 February 2013; Jeff Thomas, call with author, 28 February 2013; Patrick Cullom, interview by author, 4 March 2013.
77. Glynn A. Edwards, "Enigma of Email" (presentation at Rare Books and Manuscripts Conference, San Diego, CA, June 2012), 3, <https://docs.google.com/open?id=0B35ZUUIDskPhWGFYdTFjcDJCYVE>.
78. The processing plan for the Stephen Jay Gould papers can be seen in AIMS, 99-103. The Processing Workflow can be seen on pages 120-23.
79. Jeff Thomas, "What Do I Do With The Black Box? Processing the Electronic Records in the Deborah Pryce Papers," *Congressional Papers Roundtable Newsletter*, (Winter 2013), 24, [http://www2.archivists.org/sites/all/files/2013\\_Winter.pdf](http://www2.archivists.org/sites/all/files/2013_Winter.pdf).
80. Seth Shaw, the Electronic Records Archivist at Duke University, is working on a prototype of a system to provide virtual access to records. Given the success of the

- Duke Data Accessioner that he developed, this is an promising initiative. Interview by author, 7 March 2013.
81. "Guidelines for Creating Agreements at Stanford University," in AIMS, 117-19. More general guidelines for a collection development policy can be found on page 6; the key elements of a donor agreement are listed on pages 9-10.
  82. Thomas, 24.
  83. Adam Bergfeld, call with author, 11 March 2013; Sarah Ticer, email message to author, 15 March 2013.
  84. For more information on FOIA, the Clinton Library has a web page that explains its applications to presidential records. See <http://www.clintonlibrary.gov/foia.html>.
  85. Given the fact that these responses are more reflective of being new to handling born-digital materials than they are informative about a particular repository, the author chose to anonymise these results.
  86. Anne Gilliland-Swetland, "Digital Communications: Documentary Opportunities Not to Be Missed," in *American Archival Studies: Readings in Theory and Practice*, ed. Randall C. Jimerson (Chicago: Society of American Archivists, 2000), 589-90.
  87. Goldman, 11.
  88. Mark. A. Greene, "The Power of Archives: Archivists' Values and Value in the Postmodern Age," *American Archivist* 72 (Spring/Summer 2009): 18.
  89. Mark. A. Greene, "The Power of Meaning: The Archival Mission in the Postmodern Age," *American Archivist* 65 (Spring/Summer 2002): 50.
  90. Cunningham, "Archival Management of Personal Records," 99. Even more colorfully, Cunningham explains in a later article that the Australian phrase for postponing a decision is "putting them on the never-never." Adrian Cunningham, "Waiting for the Ghost Train," 56.
  91. Patrick Cullom, interview by author, 4 March 2013.
  92. Cunningham, "Waiting for the Ghost Train," 63.
  93. Ian G. Anderson, "Are You Being Served? Historians and the Search for Primary Sources," *Archivaria* 58 (Fall 2004): 83.

94. Kirschenbaum, Ovenden, and Redwine, 62-64.
95. Christopher A. Lee, et al., "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions," *D-Lib Magazine* 18 (May/June 2012), <http://www.dlib.org/dlib/may12/lee/05lee.html>.
96. "Campus Case Studies," Society of American Archivists, accessed 28 February 2013, <http://www2.archivists.org/publications/epubs/Campus-Case-Studies>.
97. Bantin, 30.
98. For example, the Data Archiving Policy of the National Science Foundation can be viewed at <http://www.nsf.gov/sbe/ses/common/archive.jsp>. The journal *Nature* also requires that data and materials be made public before articles will be published. See <http://www.nature.com/authors/policies/availability.html>.
99. Neil Beagrie, Robert Beagrie, and Ian Rowlands, "Research Data Preservation and Access: The Views of Researchers," *Ariadne* issue 60 (2009), <http://www.ariadne.ac.uk/issue60/beagrie-et-al>.
100. Richard Pearce-Moses, "Janus in Cyberspace: Archives on the Threshold of a Digital Era," *American Archivist* 70 (Spring/Summer 2007): 22.
101. Veerle Van den Eynden et al., "Managing and Sharing Data," Colchester, UK Data Archive, 2011, 35, <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>.
102. Digital Curation Centre, "Resources for digital curators," <http://www.dcc.ac.uk/resources>.
103. Reference Model for an Open Archival Information System (OAIS), June 2012, <http://public.ccsds.org/publications/archive/650x0m2.pdf>.
104. The Section 108 Study Group Report, March 2008, <http://www.section108.gov/docs/Sec108StudyGroupReport.pdf>.
105. *Trustworthy Repositories: Audit and Certification*, February 2007, [http://www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf).
106. Cal Lee, "A Talk on Digital Preservation," Information in Life Digital Video Series, University of North

- Carolina at Chapel Hill, YouTube, 44:41, May 31, 2007, posted by UNC-Chapel Hill, October 7, 2007, [http://www.youtube.com/watch?v=RHy9CW\\_vMp4](http://www.youtube.com/watch?v=RHy9CW_vMp4).
107. Nelson et al., 19.
  108. Interview by author, 14 February 2013.
  109. Hedstrom, "Electronic Records Program Strategies," 6.
  110. Liz Bishoff, "Digital Preservation Assessment: Readyng Cultural Heritage Institutions for Digital Preservation" (paper, DigCCurr2007: An International Symposium in Digital Curation at the School of Information and Library Science at the University of North Carolina at Chapel Hill, April 20, 2007), 2, [http://ils.unc.edu/digccurr2007/papers/bishoff\\_paper\\_8-3.pdf](http://ils.unc.edu/digccurr2007/papers/bishoff_paper_8-3.pdf).
  111. Greene, interview by author, 18 February 2013.
  112. Nelson et al., 77.



Copyright of Journal for the Society of North Carolina Archivists is the property of Society of North Carolina Archivists and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

## REVIEWS

**Mary A. Caldera and Kathryn M. Neal, eds. *[Through the] Archival Looking Glass: A Reader on Diversity and Inclusion*. Chicago: Society of American Archivists, 2014. 296p. Appendices, index, notes, and illustrations. \$69.95 (nonmember); \$49.95 (member).**

Diversity is a tricky concept to define. The subjectivity and variance inherent in the word itself weights any explanations, advice, or suggestions for its implementation with a nuanced, sometimes unhappy history of bold words, imbalances of power, and admirable expectations that go unrealized. The archival world, even in its tendency toward liberality of politics and spirit, is in no way immune to the ills of exclusion and shortsightedness. To that end, *[Through the] Archival Looking Glass*, edited by Mary A. Caldera and Kathryn M. Neal, contains ten essays by a non-homogenous collection of thinkers and doers who use a blend of narrative and analytical writing to advocate for a wide range of methods that engender and encourage diversity in the archives. The essays stand on their own quite easily (though they have a greater impact, of course, in their collected form), making them ideal to share with students or mentees, as well as colleagues looking to develop new policies. The writing style sometimes becomes a bit chock-full of references and dates, but there is also much strength to be found in the deliberate and faithful attention paid to the history being discussed. While much shop is talked in the course of these pages, there are also relatable stories being told and clear questions being posed, which makes the book

appropriate for those entering the profession as well as those looking to make a change from within the field.

The editors' thoughtful introduction provides an overview of existing archival scholarship about diversity (a wonderful starter bibliography for anyone interested in delving deep), while setting forth the book's mission quite simply: "our purpose is neither to define diversity in archives nor to prescribe ways to achieve it... Instead, our desire is to illustrate the multitude of perspectives and issues, to provide a vehicle by which new voices can be heard along with more familiar ones and new concepts examined along with new treatments of established ideas" (xix). The argument that diversity cannot be attained through one overarching solution is crucial for deconstructing readers' expectations for answers. However, the sometimes uncomfortable stories of truth, conflict, and vulnerability presented in this collection come together as a possible approach to begin demystifying the complexity of the issue.

Right out of the gate, two essays ground the book's exploration of diversity in the personal. Valerie Love and Marisol Ramos delve deeply into their own experiences to present "Identity and Inclusion in the Archives: Challenges of Documenting One's Own Community." Both authors discuss the ways in which archivists who represent a specific cultural community are often saddled with outsized expectations in terms of increasing diversity, instead of archives establishing more wide-scale efforts to make diversity an institutional goal for all of its employees. Love and Ramos incorporate their personal narratives with well-supported arguments for redefining archival practice, e.g. more collaborative and participatory documentation/arrangement/description procedures, developing trust with

local and underrepresented communities, and straightforward acknowledgement of limitations and historical biases at work. The very next chapter – Mark Greene's essay “Into the Deep End: One Archivist's Struggles with Diversity, Community, Collaboration, and Their Implications for Our Profession” – calls on archivists to “reconsider their traditions” (41) by persistently questioning who can and should control archival material, by sharing professional expertise with community archives and archivists, and by engaging with the concept of diversity on a multiplicity of levels instead of just well-meaning words. Both essays strike a balance of honesty and optimism, even while communicating a complicated history of struggle. It is a disarming, important way to begin a book examining professional attitudes and practices. Accepting these individual voices as vital catalysts of change is key to making archivists embrace not only the immensity of their own personal histories, but to open their perspectives to the existence of so many other histories as yet undocumented by the archive.

The case studies that populate a large portion of the book hereafter are helpful and enlightening examinations of archives and archivists that moved beyond traditions of collection development, notions of institutional supremacy and authority, and even the definitions of diversity set forth by Society of American Archivists itself over the course of the past several decades. T-Kay Sangwand's essay “Revolutionizing the Archival Record through Rap: Cuban Hip Hop and Its Implications for Reorienting the Archival Paradigm” begins with a section entitled “Interrogating Diversity,” a phrase that archivists shouldn't shy away from in their own practice. Sangwand's writing succeeds in

questioning the efficacy of traditional archival practices, particularly in regards to sometimes ephemeral or intangible records and “oral, aural, and kinetic forms of memory” (101). Situating the sustainability of our profession in relation to the evolution and inclusiveness of its ideologies is a critically important point that Sangwand makes elegantly while discussing Cuban Hip Hop’s artistic and socio-political origins. Sangwand also notes that archivists must “reorient their role from custodians of records to facilitators of preservation of records” (103) – a concept that could help reinvent archival approaches to acquisition, stewardship, and outreach.

Several other case studies function as in-depth examinations of the intersections between historical trends and events with current archival practice. Sonia Yaco and Beatriz Betancourt Hardy’s essay “A Documentation Case Study: The Desegregation of Virginia Education (DOVE) Project” gives a thorough, step-by-step procedural on the implementation of documentation strategies and collaborative efforts for a large-scale cataloging project. “Archives (Re)Imagined Elsewhere: Asian American Community-based Archival Organizations” by Vivian Wong, Tom Ikeda, Ellen-Rae Cachola, and Florante Peter Ibanez, sketches out a vibrant, detailed description of a specific community’s evolving history-keeping practices, citing specific examples such as the Filipino American Library and Densho: the Japanese American Legacy Project. Wong and her collaborators rightly explore the possible tension between local archival efforts and that of institutional repositories, along with a thoughtful deconstruction of the term “Diaspora” that speaks to how archives, like so many cultural communities, are often caught between the past and

present, tradition and innovation. Jeffrey Mifflin explores this rift in his essay “Regarding Indigenous Knowledge in Archives,” where he examines the preservation of oral tradition in an increasingly digital world and the complex privacy concerns regarding relocation of Indigenous materials into non-Indigenous repositories. In her essay, “Respecting Their World: How the Braun Research Library Works with Native Communities,” Kim Walters explores incorporating Native American cultural materials into collections and the necessity for a constant dialogue with the creators of these histories. Having access to lucid explanations of how others have navigated such complicated seas is invaluable and instructive.

*[Through the] Archival Looking Glass* also includes essays by Anne Gilliland, Daniel Hartwig and Christine Weideman, and Sharon Tibodeau that focus on improvements that can be made in outreach and education about how archives function, the breaking down of institutional barriers preventing diversity, and the necessity for shaping new generations of archivists who will be able to confront personal and professional biases by exercising pluralistic methodology. The paths suggested toward cultivating a culture of archival inclusiveness are not simplistic by any stretch of the imagination. Gilliland stresses in her conclusion that “It is both an ethical imperative and pragmatically to everyone’s mutual benefit to promote that pluralism to ensure the best and most appropriate stewardship of all communities’ records and memory texts as well as the continued relevance of the archival field in a plural world” (268). And these proposed ideas about diversity, questioned and committed to by a multi-cultural collection of archival professionals, should be

advanced and elaborated upon by those in the field. Indeed these ideas should be proposed *repeatedly*, questioned *thoroughly*, and committed to *wholeheartedly*.

These essays are meticulously cited, thought-provoking, and sometimes achingly earnest. These essays talk about archivists as thoughtful, activist individuals, as people of color and LGBTQ people shouldered with the expectation of being standard-bearers for the communities they represent, and most of all, as professionals who understand that it will take a reinvention of what best practices truly are in order to make archives more fully a part of the communities and histories they purport to represent. These essays will remind the reader that much discussion has happened within the Society of American Archivists about how to make the profession as a whole more inclusive and more representative of all societal facets, but that we have miles to go before we sleep. And finally, these essays will patiently explain that talking about change isn't the same as enacting change, no matter how good intentions might seem. There is work to be done and really, there always will be.

*Rosemary K. J. Davis*  
*Amherst College*

**Beth C. Thomsett-Scott, ed. *Marketing with Social Media*. Chicago: American Library Association, 2014. 178 p. Contributors, index. \$76.00 non-member (Print book/eBook bundle); \$68.40 member (Print book/eBook bundle).**

*Marketing with Social Media*, edited by Beth C. Thomsett-Scott, is a comprehensive overview of different

Copyright of Journal for the Society of North Carolina Archivists is the property of Society of North Carolina Archivists and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.



advanced and elaborated upon by those in the field. Indeed these ideas should be proposed *repeatedly*, questioned *thoroughly*, and committed to *wholeheartedly*.

These essays are meticulously cited, thought-provoking, and sometimes aching earnest. These essays talk about archivists as thoughtful, activist individuals, as people of color and LGBTQ people shouldered with the expectation of being standard-bearers for the communities they represent, and most of all, as professionals who understand that it will take a reinvention of what best practices truly are in order to make archives more fully a part of the communities and histories they purport to represent. These essays will remind the reader that much discussion has happened within the Society of American Archivists about how to make the profession as a whole more inclusive and more representative of all societal facets, but that we have miles to go before we sleep. And finally, these essays will patiently explain that talking about change isn't the same as enacting change, no matter how good intentions might seem. There is work to be done and really, there always will be.

*Rosemary K. J. Davis*  
*Amherst College*

**Beth C. Thomsett-Scott, ed. *Marketing with Social Media*. Chicago: American Library Association, 2014. 178 p. Contributors, index. \$76.00 non-member (Print book/eBook bundle); \$68.40 member (Print book/eBook bundle).**

*Marketing with Social Media*, edited by Beth C. Thomsett-Scott, is a comprehensive overview of different

social media tools that librarians can consider when marketing their institutions. This guide, published by the American Library Association (ALA), focuses on how libraries can use these technologies but contains very few mentions of archives and special collections. In the preface, the author states that one of the book's goals is to "address multiple technologies or look at implementing a technology in a step-by-step manner" (ix). This goal was accomplished throughout the publication for the audience of librarians; however, archivists may be disappointed with the lack of case studies from archives and special collections.

The first chapter of *Marketing with Social Media* is an introduction that provides a sufficient overview of each topic, introducing the specific social media tools that will be detailed throughout the book and accurately describing what the reader should expect in the remaining chapters. Additionally, the introduction does an excellent job of dispelling the common misperception that an institution should have a presence on every social media site. Instead, it asserts that "Librarians must necessarily be selective about which sites to use, as the time and effort involved in creating a presence on every available site far outweighs the potential rewards" (3).

*Marketing with Social Media* is extremely easy to follow, with each chapter detailing a specific social media technology: Facebook, wikis, video-sharing services, Pinterest, Google+, Foursquare, blogs, QR Codes, and Twitter. However, this book may be more helpful to a user looking for information on one specific tool as opposed to those needing a comprehensive guide. Since each chapter provides guidelines for implementation of one social media instrument and general guidelines for these are similar, some

of the information provided is repetitive. For instance, there are several recommendations several times throughout the book that a staff member's personal email address not be used for the username for a site, so other staff members have access to the account in case that person leaves the institution. Using this publication as a reference tool when considering implementing a new social media initiative would be a better use than reading it from cover to cover as some information, such as the username example, is frequently duplicated.

A major theme that carries throughout the entire publication is how institutions can evaluate the impact and success of using social media. A point of emphasis in several chapters is that evaluation for success depends on the social media application. In the chapter about Twitter, authors Laura Carscaddon and Kimberly Chapman explain that to evaluate the success of a library's Twitter account is more than keeping track of the number of followers a library receives; who the library interacts with and how the library is mentioned among other users on Twitter should also be taken into consideration when evaluating the success of social media sites. This is an excellent example that underscores the complexity of evaluating success in using social media both quantitatively and qualitatively.

This book is designed for beginners at libraries who are not as experienced with social media tools. The overview for each social media platform detailed is a great resource for these readers. Additionally, each chapter gives a brief history of the technology and what demographic generally uses it, which is extremely useful information for librarians deciding which technologies to use to market different library services to various age groups. Chapters also list steps for

implementation, security considerations, and the pros and cons of using each social media tool. For example, Amy West writes that Google+ may not be the best way to “reach the largest number of users on the most regular basis,” and therefore “currently there are a few examples of librarians with active presence and engaged users in Google+” (83). This chapter is one of the most helpful because the author acknowledges the downfalls of the site while still presenting excellent reasons for librarians to monitor it for possible future implementation at their institutions.

There are many social media tools that information professionals can choose to help market their institutions, and this book serves as a good starting point with specific examples of different implementations. The danger in writing a book about technology, and specifically social media applications, is that it is a constantly changing topic. In a few years these platforms may be obsolete, with other tools rising to the forefront of the industry. *Marketing with Social Media* is a helpful publication especially for archivists who also serve as librarians, but will likely have a limited period of usefulness before the information it provides and technologies it describes are outdated.

*Pam Richter*  
*Baldwin Borough Public Library*

Copyright of Journal for the Society of North Carolina Archivists is the property of Society of North Carolina Archivists and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

**Abby Clobridge. *Building a Digital Repository Program with Limited Resources*. Oxford: Chandos Publishing, 2010. 239p. Figures, appendices, bibliography, and index. \$85 (print); \$85 (ebook); \$170 (print & ebook).**

Intended for a broad audience spanning on-the-ground practitioners as well as library administrators, *Building a Digital Repository Program with Limited Resources* provides a strong starting place for getting a digital repository program off the ground, including institutional repositories and repositories of digital special collections. For institutions that already have a program established, this book is intended to aid in the ongoing assessment and re-tooling of an existing program. Published in 2010, it still provides a useful roadmap for smaller to mid-sized institutions that are implementing or evaluating a digital repository program in 2014.

The book's central principle holds that program development encompasses a broader range of activities than simply technical implementation. Drawing on her experience at Bucknell University and the Harvard Kennedy School of Government, Clobridge recognizes that building a repository is not just about software, file formats, and metadata; building a repository is also about defining an audience, identifying partners, and marketing collections. Moreover, each of these activities must be conducted with a strong awareness of institutional context.

Part One provides an overview of how to launch a digital repository program, including chapters on strategic planning, technical environments, and staffing roles. Extremely helpful sample documents in the chapter on strategic planning include:

- spreadsheets designed to audit internal resources of staffing, hardware, and software;
- needs assessment worksheets;
- repository mission and vision documents; and
- worksheets for relating university-wide strategic planning goals to repository action items.

The chapter on staffing includes sample position descriptions for full-time repository personnel as well as learning objectives for interns. Geared toward building institutional consensus and clarifying needs and goals, these sample documents enable new repository coordinators to hit the ground running.

Part Two shifts its focus from launching a new digital repository program to sustaining one that has already been established. Chapter topics include metadata, project management, acquiring and marketing content, open access, long-term sustainability, assessment, and incorporating Web 2.0 elements. After a brief overview of metadata in general, the chapter on metadata emphasizes decision-making (e.g. questions to ask when selecting a metadata standard) and documentation (e.g. defining a data dictionary for the repository or for a collection). The chapter on project development includes a sample digital project proposal form – an extraordinarily helpful exercise for digital projects proposed by library personnel as well as those proposed by faculty and other campus partners. In addition, this chapter emphasizes the importance of batch processes in any production environment and points to Microsoft Excel and command-line processes for optimizing workflows. Like Part One, Part Two focuses on both the nitty-gritty and the

organizational aspects of setting up a digital repository program for success and sustainability.

The book is organized such that it could be read cover to cover or quickly consulted as needed. The chapters often represent discrete stages of planning or implementation, and each chapter includes guiding principles and references for further reading. Clobridge's prose is cogent and concise. The figures often consist of sample documents that vividly illustrate planning activities.

Since 2010, some trends that were not covered in-depth in this book have come to the forefront. Chief among these are research data management and the digital humanities, both of which would now be core considerations of launching or re-tooling a digital repository program for many institutions. Additionally, since 2010, more tools have been added to the digital repository assessment toolkit. These include ISO 16363: Audit and Certification of Trustworthy Digital Repositories and the National Digital Stewardship Association's Levels of Preservation project. Even with these recent developments, Clobridge's *Building a Digital Repository Program with Limited Resources* remains a pragmatic and holistic guide for both wide-angle institutional planning and day-to-day management activities.

*Chelcie Juliet Rowell*  
*Wake Forest University*



Copyright of Journal for the Society of North Carolina Archivists is the property of Society of North Carolina Archivists and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.