Digitized Special Collections and Multiple User Groups

Gretchen Gueguen
Head, Digital Collections
East Carolina University
Greenville, North Carolina
(252) 328-4978 TEL
guegueng@ecu.edu E-MAIL

Digitized Special Collections and Multiple User Groups

Abstract

Many organizations have evolved since their early attempts to mount digital exhibits on the web and are experimenting with ways to increase the scale of their digitized collections by utilizing archival finding aid description rather than resource-intensive collections and exhibits. This paper examines usability research to predict how such systems might effectively be used and highlights a digital library and finding aid system that utilizes a single repository of digitized objects to fuel two types of user-discovery systems: a typical digital collections interface with item-level access and a finding aid that incorporates digitized items at the aggregate level.

Introduction

While librarians have debated and discussed the digitization and full-text searching of books for some time, a recent surge of interest has developed in the archival community on a more elusive target: large-scale digitization of special collections and archival materials. Many organizations have moved beyond their early attempts to mount heavily contextualized exhibits and thematic collections on the web and are experimenting with ways to increase scale without overwhelming traditional work. However, a tension exists between many digitization initiatives that create digital collections with robust item-level metadata and context, and the traditional archival manner of defining and describing collections at the aggregate or group level.¹

At the same time, questions have arisen about how these differing methods of access affect the success of different user groups in finding and using digitized primary sources. Will novice users be confused when encountering the aggregate-level description of the finding aid, expecting to encounter the traditional item-level description they have previously seen? Will researchers used to the finding aid be able to use the item-level collections efficiently?

The tension between these conflicting paradigms led the staff of the Digital Collections department at East Carolina University to investigate the typical users of the special collections materials they were digitizing to develop more suitable means of access for digitized materials on the web. This paper will examine the creation of a digital library and finding aid system that supports multiple access interfaces to suit the needs of two user groups simultaneously: undergraduate students and humanities researchers.

Humanities Users and Research

The design of any service or system often begins with an identification of the expected audience. It is true, of course, that on the web any user may find, and find useful, the materials digitized by archives and special collections. However, one of the strongest indicators of whom the collection may be useful *for* are the traditional users of the physical resources that are being digitized: researchers in the humanities.

Research on the information seeking and usage behaviors of scholars in the humanities has been undertaken in many contexts: archival research, library database and catalog searches, writing practices, and teaching practices. Humanities users have been shown to favor a research style that is initially broad and benefit from browsing a large and diverse set of resources. Searches often involve retrieving large sets of results, either by browsing or using multi-term searches, and sorting through sets for items of interest. This allows the scholars to serendipitously retrieve records that meet their specific, though perhaps unarticulated needs, while keeping the possibilities open for potentially overlooked or unconventional sources².

With advances in technology that increase the scope of what the scholar can search and how they can search, humanities researchers continue to rely heavily on texts and engage with them deeply throughout their research process: reading and rereading, annotating, collecting and organizing ³. The careful reading of texts in turn leads to the reading of other texts that are referenced therein⁴.

What has not been studied as extensively are the habits of experienced researchers and their discovery and use of primary sources when they are available on the web, particularly in non-book formats. An exception to this is the final report for the grant project "Extending the Reach of Southern Sources." The project involved extensive outreach to historians and other scholars interested in the Southern History Collection at the University of North Carolina at Chapel Hill. In particular, the scholars were asked how

they would like the library to approach the mass digitization of archival materials. Their results suggest that most scholars embraced the opportunity to do more powerful searching across documents and the convenience of the digital medium. However, the researchers showed little interest in several traditional methods of providing digitized materials on the web, such as selective digitization and exhibits with contextual and analytical material meant to guide the user. ⁵

Instead, the researchers wanted entire collections digitized, even the scraps and odds and ends that may seem unimportant or uninteresting to archivists. In addition, they wanted the digitized representation to match as closely as possible the traditional organization in the archives⁶. They wanted the research paradigm of the archives, where they can engage in their favored behaviors of browsing through entire collections to make their own determinations of what is relevant to their research question with the enhancement of being able to perform searches and browse by certain features to better target which collections and portions of the collections they needed.

Undergraduate Students and Teachable Moments

While humanities researchers have a long history of use of archives, evolving notions of the value of using primary resources in education⁷ have driven more undergraduate students into archives and special collections. In fact, numerous studies over the past ten years have emphasized the role of primary sources in undergraduate research, but most suppose that students get this exposure in person, in the archives. In 2004, Elizabeth Yakel suggested a model for a standardized archival instruction that goes beyond mere orientation, instead addressing comprehensive information literacy⁸. Others focus on the role of the archivist or librarian in special collections in helping to encourage and co-develop curriculum with faculty to include archival research⁹.

At the same time, studies of undergraduates in general find that they are extremely convenienceoriented¹⁰. When students lack the structure of an in-person archival orientation, they struggle with
finding and using primary sources and particularly archival materials. Archer, et.al. described the findings
of a usability study on a web-based primary resources tutorial meant to supplement or in some cases
even stand alone as instruction in primary sources and archival materials. Their study uncovered a
relatively high knowledge of library tools such as the catalog and databases, but little to no familiarity
with the skills needed to perform research with primary sources: knowledge of what primary sources are
and how to use archival finding aids to get access to them. ¹¹

Although Elizabeth Yakel in 2004 found that novice users had great difficulty in navigating finding aids and understanding archival jargon¹², Archer, et. al.¹³, as well as Scheir¹⁴, found that students learn archival terminology as well as the purpose and use of a finding aid relatively quickly. However, students' searching styles may still not be well suited to finding archival information. Again and again in their study, Archer et.al found that students preferred any form of guidance they could find: browsing subject lists, using research guides, and relying on the perceived relevance ranking of broad searches¹⁵. In general, the students in their study were far more likely to have received instruction in library tools and sources such as the catalog and databases that use item-level description and increasingly include direct links to digitized materials. They brought these skills with them to their approach to searches of archival materials.

Experiences with using full-text databases are also changing student expectations and archives are no exception. As Prom showed in 2004, novice searchers expected the finding aid to include digitized material rather than just serve as a guide to collections in the archive¹⁶. Furthermore, the abstract representation of a finding aid does not mesh with the types of experiences students may increasingly have at the K-12 level as described by Malkmus in her overview of the literature surrounding the use of

primary sources in these settings. Malkmus showed that many K-12 classroom settings are document-focused and deal with archival materials at the individual item-level¹⁷.

Though not favored by the experienced researcher, the digital exhibit is a good model for helping students to interact with primary resources. Exhibits and collections, because they offer interpretive content and showcase only carefully selected materials, offer the student guidance through a topic and can frame the resources within a historical context. Exhibits can also help to focus attention on documents important to telling a story or making a pedagogical point. Too many documents can be overwhelming and lessen the impact. Helping students achieve focus is even more important in webbased research where the scope of documents available is exponentially increased. The structure of digital exhibits, with item-level description and a focused sampling of materials, are therefore more likely to be readily understood by undergraduates who have not received an archival education.

A Multi-Purpose Repository

With these studies in mind, the Digital Collections department at East Carolina University's J.Y. Joyner Library endeavored to design a digital repository that could meet as many needs of these two divergent groups' as possible. Work on the project began with the need for a Digital Asset Management System (DAMS) at Joyner. For many years, the Digital Collections department only created digital exhibits or stand-alone collections. A researcher may have been able to perform searches within a given exhibit or collection but could not search across these resources, much less make the kinds of serendipitous and tangential kinds of discoveries between resources in different collections that they rely on to do their work.

In practical terms, such exhibits are also resource-intensive to mount. Careful selection processes, writing contextual essays, and designing websites with aesthetics and navigation that highlight the

materials all lead to significant development time. In order to increase the digitization of special collections material significantly more resources need to be devoted to their creation or this context must be stripped away.

When Digital Collections began to design their DAMS, the goal was to increase the scale and cross-searchability of all of the digital projects as well as to create a single system to lower the costs of continually creating and maintaining new resources. Built using the Ixiasoft's TEXTML server software, the repository is a native XML storage/retrieval database populated with METS XML records. Each digitized object (text, image, audio, video, or some combination of these) is described in a METS record (containing a mixture of appropriate metadata schemas such as MIX, MODS, TEI, and DC) with links to the digitized content. TEXTML then indexes these records and provides results in the web-based asset management system built on the ASP.NET framework.

The final DAMS interface was completed in February 2009 ¹⁸. The public web access point was simply named *Joyner Library Digital Collections* to emphasize that this resource provided access to a multitude of materials rather than being focused on a specific theme.

Although the repository has many similarities to the traditional digital collection or exhibit (item-level orientation, heterogeneous subject-based collections), specific features are included to facilitate powerful searching, browsing and sorting behaviors that facilitate the types of searching done by both the undergraduate and humanities scholar audience. Some of these specific features included:

Broad thematic collections allow casual or novice users a guide to the scope of the repository.
 By organizing around generic themes such as African American History, Transportation, and
 Politics and Government, rather than institutional collections, the undergraduate is provided with some context and guidance through the content.

- Collection templates provide some limited ability to contextualize with space for brief introductions to collection, while maintaining an institutional look and easy universal navigation.
- **Subject cloud visualization** is another guidance tool for novice users that also provides advanced users the opportunity to quickly gauge the scope of the collection.
- Hypertext links in records are created for subjects and collection names that initiate a search through the entire repository for other objects similarly indexed. These links are critical to exploration strategies for more advanced users as well as novices or those who found repository records through an open internet search. These hypertext links are also created for relationships within the archival organization where appropriate ("See other items in this folder," "...in this box", etc.) for advanced users (figure 1).

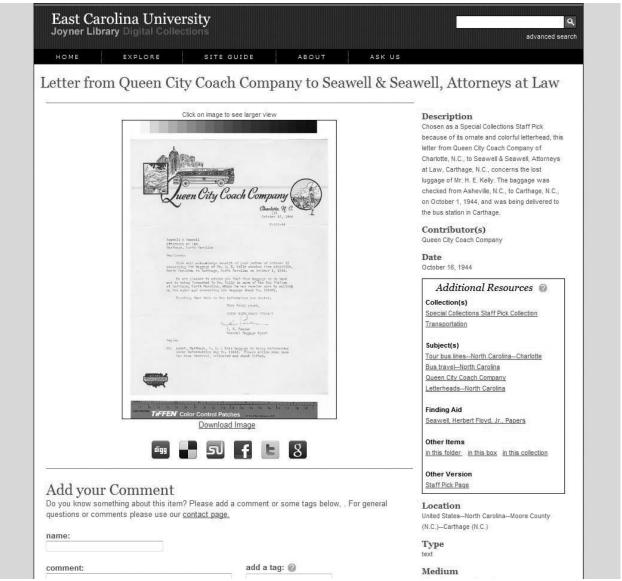


Figure 1: An example of a metadata record from *Joyner Library Digital Collections* showing hypertext links in the "Additional Resources" section.

• Facets for refining searches: a powerful tool for narrowing broad initial search results which are favored by both groups (figure 2).

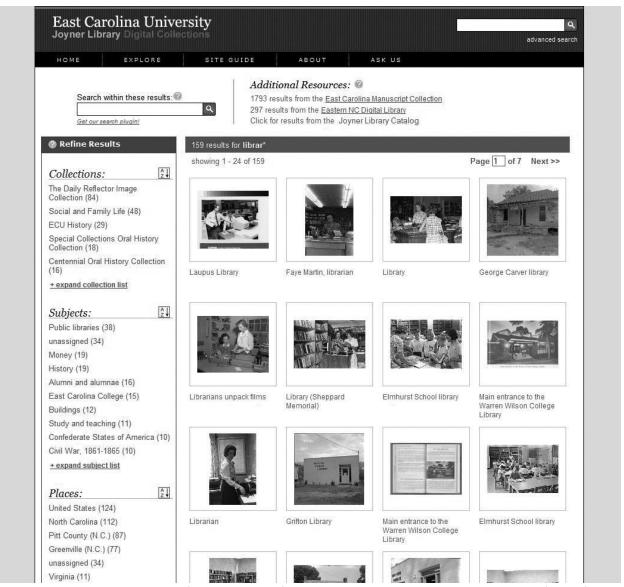


Figure 2: A typical results page from a search of *Joyner Library Digital Collections* showing facets for refining searches.

Commenting and tagging records provides another layer of context in some cases, particularly
since the digital collection contains many images of local history which have been annotated
extensively by the local community with substantive information.

An Interface for Research

Many of the features in the Digital Collections interface are not distracting to a serious researcher in the humanities, and in fact some of the features like hyperlinks within records based on archival organization were created specifically for this group. However, overall the repository supports the novice or casual user more than the serious researcher. Designing a system that really supports humanities scholars' research styles requires paying more attention the traditional tool for accessing archiving material: the finding aid.

The EAD finding aids for the library's East Carolina Manuscript Collection are composed of more than 1800 individual collections. These finding aids were originally encoded and placed on the web in 2002. The full-text EAD XML finding aids are stored in a separate TEXTML index and rendered through an XSL stylesheet. Links from the EAD to digitized objects in the DAMS have been added on a periodic basis over the years. These references were placed in <did> elements within a <c04> element for the individual item. The <did> included the reference to the object's persistent identifier, or pid, which the stylesheet used to create the hyperlink. In addition, an item title and date were added. This configuration allowed the title and date to be searchable and to display within the container list and the link to be to the discrete item:

All of the information in the <did> element had to be added by hand. The process of maintaining this practice was difficult as literally hundreds of items could be added to the repository in a week and a serious backlog problem was developing. Furthermore, should any change be made to update the item's title or date, that information had to be updated in two places, both the DAMS record and the EAD.

The user experience suffered as well — each individual link could be clicked on one at a time, but there was no way to view the materials together as a collection. The links also led back to the DAMS and its home interface, breaking away from the finding aid look and feel and entering a different domain.

Based on the growing backlog problem, the dissatisfaction with the interface, the results of the study by the Southern Historical Collection¹⁹, as well the interface created by The Archives of American Art's *Collections Online*²⁰, Digital Collections decided a redesign of the finding aids was in order that would better incorporate digitized materials by using the backbone of the DAMS. The Text and Markup Coordinator tackled the stylesheet redesign, and worked with the Digital Library Developer to implement navigation under the direction of the Digital Initiatives Librarian. The redesigned *Joyner Library Collection Guides*²¹ debuted April 12, 2010.

The interaction with the digitized objects occurs in one of two ways. First, the finding aid contains three "tabbed" views. The code for the header and the tabs in contained in an ASP.NET Master page. The first tab, entitled "About this Collection," renders the full text of the finding aid itself in an XSL stylesheet (figure 3). The second tab, "Add/View Notes," contains an area for the addition of notes and commentary

by users or archivists. The third tab, called "View Digitized Objects," contains digitized objects from the collection (figure 4).

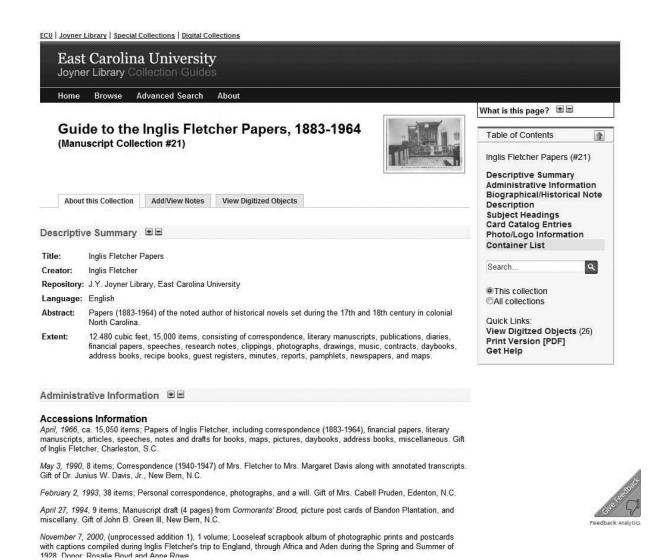


Figure 3: The "About this Collection" tab in the revamped EAD stylesheet.

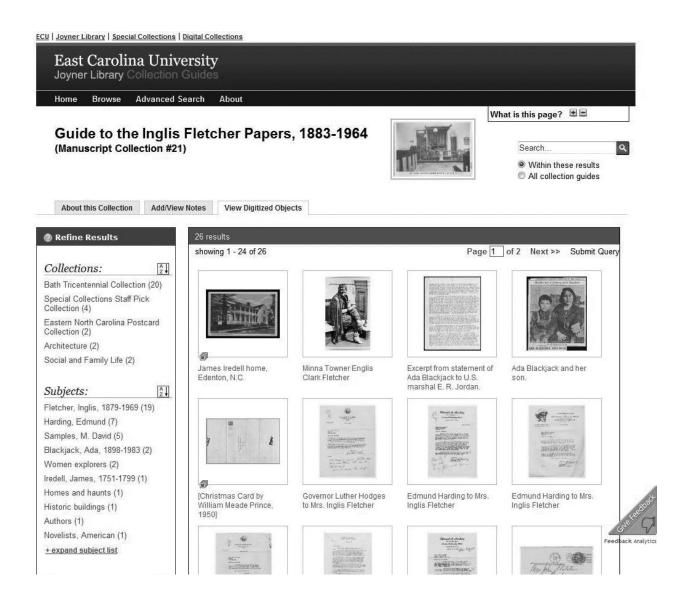


Figure 4: The digitized objects "tab" in the revamped EAD stylesheet.

The internal markup of the EAD is much the same, with the addition of a <dao> element which also references the pid:

The <cO4> and <unittitle> elements are still included in order to have that information available for searching within the EAD record. The addition of the <dao> element was included in this redesign since it is a more standard container for this type of information. It can be easily updated with the root URL to create a valid link should these records be shared in the future. Although, at the moment, this information is redundant with the <unitid type="pid">, this redundancy is trivial and may well be removed in the future.

The biggest difference between the old and new encoding was the way in which it is created. The process of adding this information has been folded into the digitization process and partially automated. First, the archival location number is recorded in the MODS metadata record in the <identifier> element when the digitized item is added to the repository. This information is recorded in an alphanumeric string such as " 741.21.d.6" where "741" is the collection number, "21" is the box number, "d" is the folder, and "6" is the item number (usually generated as a product of digitization, this number is rarely assigned in the original processing). Next, the digitization operator (usually a student employee) selects the object's collection from a list on a specially created web form. The form then sends a runs a routine that searches the repository for members of that collection and updates the container list with the new <c04> elements containing <did> and <dao> elements populated with information from the DAMS. Any subsequent changes to the item's title or date in the DAMS will also be reindexed into the EAD as part of

the metadata workflow. In this way, the extra information contained in the title and date is retained by creating the simple pid-based link, but the consistency problem between the two resources is solved.

Links are still included in the container list at the item level to the direct URL of each object, generated from the pid in the <dao> element. However, with the consistent formatting of the archival location in the <identifier> element of the MODS records, links can also be created to do a search of the DAMS for items belonging to a container such as "741.21" or "741.21.d." This means that, in addition to links at the item level as before, entire groups of materials can be linked in a digital simulation of opening the entire folder or box in the archival setting.

Within the interface, clicking on a link at a folder or box level or clicking the "View Digitized Objects" tab sends a request to the DAMS for all objects that belong to the container or collection. The XML data returned by the DAMS is rendered in a stylesheet that is an exact copy of the stylesheet found in the *Joyner Library Digital Collections* interface. This stylesheet is nested within the master page code for the header and tabs. Once a user clicks on and individual item within these results they are take to the DAMS record including user comments, tags, and links to other DAMS resources.

Essentially the DAMS XML database can be called from two different application layers: one in *Joyner Library Digital Collections* and one in the EAD interface. This allows the user to navigate between content in the EAD database and the DAMS without perceiving any jump from one resource to another until they view an individual item.

Evaluation

Admittedly, the system does not have several key components identified by the Southern Historical Collection or seen in *Collections Online* such as integration of description at the folder-level and the ability to leaf through or compare objects side by side. Most importantly, when the collection is not completely digitized there is no indication of which materials are not digitized in the "View Digitized

Objects" tab. This feature was highly regarded by the scholars in the Southern Historical Collection's study²².

Future usability tests, planned for this spring will evaluate the need for these features as well as others. In particular, the degree to which users do and do not need to remain within the context of the EAD master page will be evaluated. For example, is the user lost when jumping out to the DAMS to view the individual item record? Are the features built into that record, such as the hyperlinks, useful in this context? Should the "View Digitized Objects" tabs continue to mimic the *Joyner Library Digital Collections* stylesheet or would it be less confusing for it to have a more distinctive style? From the results of these tests, staff will try to determine if more work needs to be done to integrate the item level display in to the EAD master page as well. Another issue to be further investigated is the integration of folder-level description. Some preliminary attempts have been made to create folder-level objects in the DAMS rather than describe each folder member individually. It is unclear at this time if mixing itemand folder-level descriptions within the DAMS is problematic for users.

While the resources in this new version of the finding aid do not integrate the digitized object into the finding aid navigation as completely as the example of the *Collections Online*, they offer a reasonable compromise between having to build and maintain two separate systems. By facilitating features in each resource that facilitate each group's research style the impact of these resources is maximized. While the DAMS interface may be geared to use by undergraduates or other novice users, advanced researchers can still take advantage of the ability to search across a broad array of resources and harness the power of linkages in metadata that enhance serendipity and browsing. On the other hand, the finding aids allow serious researchers to search across and use archival materials in a more sophisticated way, while the incorporation of digitized materials into finding aids assists undergraduate users who may be working with them for the first time.

The tools created by East Carolina University to integrate digitized objects with EAD are in many ways specific to its own workflows, practices, traditions, and previous digital products. However, any institution that is able to engineer some flexibility between their DAMS and EAD can use some of the concepts discussed here to begin to develop their own method of integration, such as utilizing archival location numbers in identifiers in the DAMS as well as regenerating DAMS output into EAD stylesheets. Reusing the same digital objects in both an item-based digital library search platform and a finding-aid style navigation tool is an efficient and sustainable option for the library as it means creating and maintaining only one digitized object. More importantly though, it also suits different information needs and search styles. Ultimately, this serves more users and greatly increases the audience for special collections.

Notes:

¹Sterling Coleman Jr. "The Archival and Library Viewpoints of a Collection in a Digital Environment: Is There Any Room for Compromise?" *Journal of Archival Organization* 2, no.1 (2004):105.

²For further insight into studies of humanities scholars' searching practices, see: Debora Shaw, "Bibliographic Database Searching by Graduate Students in Language and Literature: Search Strategies, System Interfaces, and Relevance Judgments." *Library and Information Science Research* 17(1995): 327-345; Bates, Marcia J., Deborah N. Wilde, and Susan L. Siegfried, "An Analysis of Search Terminology Used by Humanities Scholars: The Getty Online Searching Project Report Number 1." *The Library Quarterly* 63 (1993): 1-39; William S. Brockman, Laura Neumann, Carole L. Palmer, and Tonyia J. Tidline, *Scholarly Work in the Humanities and the Evolving Online Environment*. Report produced by the Council on Library and Information Resources (2001).

http://www.clir.org/pubs/reports/pub104/contents.html; Marcia J. Bates "The cascade of interactions in the digital library interface." *Information Processing and Management*, 38, no. 3 (2002): 381-400; and Clara. M. Chu, "Literary critics at work and their information needs: a research-phases model." *Library & Information Science Research* 21, no. 2 (1999): 247-273.

- ³Carole L. Palmer, Lauren C. Teffeau and Carrie M. Pirmann, Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development. Report commissioned by OCLC Research, (2009): 35. http://www.oclc.org/programs/publications/reports/2009-02.pdf.
- ⁴This was succinctly summarized in Stephen E. Wiberley and William G. Jones, "Time and Technology: A Decade-Long Look at Humanists' Use of Electronic Information Technology." *College and Research Libraries* 61 (2000): 421-31.
- ⁵Southern Historical Collection, University Library, University of North Carolina at Chapel Hill. *Extending the Reach of Southern Sources*. Final grant report for the Andrew W. Mellon Foundation. (2009): 14. http://www.lib.unc.edu/mss/archivalmassdigitization/download/extending the reach.pdf.

- ⁷See, for example, Association of College and Research Libraries, Information Literacy Competency Standards for Higher Education. American Library Association (2000). http://www.ala.org/ala/mgrps/divs/acrl/standards/standards.pdf.
- ⁸Elizabeth Yakel, "Information literacy for primary sources: creating a new paradigm for archival researcher education." *OCLC Systems & Services* 20, no. 2 (2004): 63.
- ⁹Several studies explore this link. See Matthew Reynolds and Dale Sauter, "Engaging undergraduates in special collections through English composition: Collaborating for student success." *College & Research Libraries News* 69, no. 6 (2008): 318-21; Wendy M. Duff and Joan M. Cherry, "Archival Orientation for Undergraduate Students: An Exploratory Study of Impact." *The American Archivist* 71, no. 2 (2008): 499-529 and Jennifer Brannock, "Exposing Undergraduates to Special Collections through Bibliographic Instruction." *Mississippi Libraries* 72, no. 3 (2008):54-56.
- Two studies to explore this link in-depth include: Barbara Valentine, "Undergraduate research behavior: using focus groups to generate theory [at Linfield College]." *The Journal of Academic Librarianship* 19, no.5 (1993): 300-4; Alison J. Head and Michael B. Eisenberg, "Lessons Learned: How College Students Seek Information in the Digital Age. Project Information Literacy Project Report." The Information School, University of Washington (2009). http://projectinfolit.org/pdfs/PIL ProgressReport 2 2009.pdf.
- ¹¹Joanne Archer, Ann M. Hanlon and Jennie A. Levine, "Investigating Primary Source Literacy." *Journal of Academic Librarianship* 35, no. 5, (2009): 416.
- ¹²Elizabeth Yakel. "Encoded Archival Description: Are Finding Aids Boundary Spanners or Barriers for Users?" *Journal of Archival Organization*, 2 no. 1/2 (2004): 74.

⁶ Ibid., 34.

¹³Archer, et.al., 418.

¹⁴Wendy Scheir, "First Entry: Report on a Qualitative Exploratory Study of Novice User Experience with Online Finding Aids." *Journal of Archival Organization* 3, no. 4 (2005): 78.

¹⁵Archer et. al., 416.

¹⁶Christopher Prom, "User Interactions with Electronic Finding Aids in a Controlled Setting," *American Archivist* 67, no. 2, (2004): 247.

¹⁷For a review of teaching methods using primary sources, see the section entitled "Transformations in Teaching History" in Doris J. Malkmus "Primary Source Research and the Undergraduate: A Transforming Landscape." *Journal of Archival Organization* 6, no. 1 (2008): 49-54.

¹⁸ J.Y. Joyner Library, *Joyner Library Digital Collections*. http://digital.lib.ecu.edu Accessed: 1/15/2010.

¹⁹ Southern Historical Collection, 14.

²⁰Smithsonian Institution, Archives of American Art. *Collections Online*. http://www.aaa.si.edu/collectionsonline/Accessed: 1/15/2010.

²¹ J.Y. Joyner Library, *Joyner Library Collection Guides*. http://digital.lib.ecu.edu/special/ead Accessed: 4/12/2010.

²²Southern Historical Collection, 14.