



RESEARCHING AMERICAN HISTORY; PRIMARY SOURCES ONLINE: A LIBRARIAN'S PERSPECTIVE

P. Toby Graham

Index:

Abstracts: [English](#) | [Spanish](#) |

- o .01. [Introduction](#)
- o .02. [Descriptive Information: Catalogs and Finding Aids](#)
- o .03. [Digital Collections: "Invented" and Collection-Based Archives](#)
- o .04. [Conclusion](#)
- o [.05. References](#)
- o [.06. Selected Webliography](#) on American History

.01. Introduction ([Return to Index](#))

The University of Southern Mississippi—where I work-- is located in Hattiesburg, Mississippi. With several highways and railroads intersecting there, it prides itself on being the "Hub City" of South Mississippi. While this means a great deal if you live in places such as Shubuta, Kiln, Picayune, or Eastabuchie, the moniker it is less than impressive to the world outside of Mississippi. The University holds historical collections of national interest, particularly on race relations, but most people consider Hattiesburg, Mississippi, a bit out of the way as a research destination. Fifteen to thirty researchers visit the Special Collections reading room each day, a few hundred a month. In the same month, however, there are easily eight thousand hits on just one of the Special Collections Department's three Web sites. That tells me that my job and my audience are changing. It demonstrates that there is a demand for USM's unique archival materials online, and that my institution has an opportunity, despite its distance from most potential users, to provide service to many more people than in the past.

It is clear that information technology is changing the way that scholarly research, and particularly historical research, is done. Developments in computing, telecommunications, and imaging allow libraries, archives, and museums to provide worldwide electronic access to both descriptive information--such as archival finding aids and online catalogs--and collections of digital surrogates of actual original source material. Providing these types of information to the public via the Internet, once undertaken mainly by the largest and best-funded institutions, is becoming commonplace. The historians, students, and other users who librarians and archivists serve are also increasingly computer savvy. They demand and are critical consumers

of information technology.

These developments lead to important questions. In a 1998 article, Carl Smith asks whether "serious" history can be done online [1]. He defines "serious" history as original work that is based on the best primary evidence, that is aware of other research, and that makes a group of sustained points about its subject. A part of this issue is whether serious historical *research* can be done online. This question is as important to information professionals as to historians, because the answer may determine how heavily academic institutions invest in online access to their collections. An affirmative answer is dependent on what is or can be made available in standardized formats and whether historians can access this information in a comprehensive way.

.02. Descriptive Information: Catalogs and Finding Aids ([Return to Index](#))

Descriptive data about primary source collections is the most common type of historical information found online, and at present it remains the most important. Descriptive resources communicate what information is held where, what subjects are covered, how much material is held in a collection, whether there are restrictions on use, origin of the records, and other physical and intellectual characteristics of original source material. Use of electronic descriptive information for secondary materials is already well established. The ability to search for articles, theses, and books in electronic databases like *America: History & Life* and *Historical Abstracts* and OCLC's *World Cat* is a tremendous benefit for historical researchers. The body of descriptive resources for primary materials lacks the same kind of bibliographic power. But online access to descriptive information for original sources is constantly expanding and becoming more sophisticated. Traditionally, users of archives or other repositories of primary source material have accessed collections either through catalogs or through finding aids. In the online environment, researchers continue to use these basic tools, but the improved accessibility, functionality, and potential for cooperation with other repositories provide important advantages.

The development of machine-readable cataloging for archives and manuscripts in the 1980s eventually led to the creation of the Research Libraries Information Network's Archives and Mixed Collections file (RLIN AMC), a cooperative bibliographic network for intellectual control of historical manuscript collections. RLIN AMC currently holds approximately 500,000 records contributed by archives, libraries, museums, and historical societies throughout North America. These include 21,000 catalog records contributed by the Library of Congress' National Union Catalog of Manuscript Collections (NUCMC) unit, which abandoned its print catalog in the 1990s in favor of RLIN's database. NUCMC now provides an Internet gateway to RLIN AMC allowing researchers to access the file free of charge. The OCLC bibliographic network also collects catalog records for manuscripts, and these can be accessed from most college libraries through World Cat. In addition, electronic publishing company Chadwyck Healey offers *Archives USA*, containing records for 117,000 collections--mostly derived from NUCMC--and holdings information for 5,400

repositories.

For historians, the availability of these cooperative cataloging networks provides the opportunity to search for manuscript collections across the repositories of North America. For example, a search using the NUCMC/RLIN gateway on Alabama political activist Virginia Hamilton Durr discloses that Durr's papers are held at Radcliffe College. But the search also reveals that the Alabama state archives holds a collection of Durr oral histories and that much of Durr's political correspondence is held as a part of the National Committee to Abolish the Poll Tax Records held at Emory University.

Even in the age of machine-readable cataloging, however, the creation of archival finding aids, often called inventories or registers, remains the central activity associated with description of original source material. These finding aids describe archives collections in a very different way from catalog records. The principles of archival service dictate that collections of original source material are arranged into a number of levels, such as collection, series, filing unit, and item. This is opposed to books where description is typically only at the item level. Finding aids may provide some description for each of these levels, and they also provide information on provenance (or origin) of collections. Unlike the brief catalog record, the archival finding aid can be many pages long, providing historical context for the collection and a listing of the materials in each container. This kind of detailed information greatly enhances the functionality of a collection that may be tens, hundreds, or even thousands of cubic feet.

The centrality of the finding aid to archival work led the information professions in the Internet years toward developing and applying a standard for online finding aids. Derived from a project at the University of California, Berkeley, Encoded Archival Description (EAD) has emerged as a popular choice, if not yet a national standard, for machine-readable finding aids. Like hypertext markup language, used to create documents for the World Wide Web, EAD is an application of the rules of Standard Generalized Markup Language (SGML). Unlike HTML, EAD focuses on content rather than the appearance of Internet pages. EAD—a non-proprietary, hardware and software independent standard—supports the "dream" of providing historians with universal, union access to primary source finding aids [2]. Though many repositories continue to choose HTML over EAD, both the Library of Congress and the Society of American Archivists, as well as a number of large universities, have lent their support to the further development of Encoded Archival Description.

Mechanisms for integrated access to EAD finding aids are not fully developed, but the historian does have several options. The Library of Congress provides a list of current EAD implementers and ongoing cooperative projects. The Research Libraries Group (RLG) created a cooperative network for sharing finding aids. Called "Archival Resources," the network allows complex online searching of EAD records across institutions. A user can search simultaneously the finding aids of Harvard, Duke, the Kentucky State Archives, and dozens of other repositories. Institutions do pay a fee

for access to RLG's "Archival Resources," however.

At present, the primary disadvantage of EAD finding aids is that they cannot be viewed using a standard Web browser like Netscape or Internet Explorer. Users must download "plug-in" SGML viewer software; the most common of these is Panorama by Interleaf. So, even though SGML and EAD are not proprietary, the software required to view EAD finding aids is. For this reason, most repositories that use the technology also provide HTML versions of their finding aids or use computer programs that automatically create HTML documents from SGML.

.03. Digital Collections: "Invented" and Collection-Based Archives [\(Return to Index\)](#)

In addition to descriptive information, many libraries and archives are digitizing and providing electronic access to actual collections they hold. Digitization--in the context of archives and libraries--is creating digital surrogates of cultural or historical materials through digital imaging, digital sound and video, or other means of capturing analog information in digital form. Establishing an effective digitization project involves complex issues like copyright, selection, preservation of electronic media, description of files, quality control, staffing, and file storage. The payoff is substantial, however. Because archival materials are often unique, digitization allows worldwide access to primary sources that were previously available only at a single repository.

Edward Ayers employed digital imaging technology to create *Valley of the Shadow*, an enthralling digital resource that brings many types of documents together to present a comparative view of two Civil War towns, one northern and one southern. Michael O'Malley and Roy Rosenzweig call this type of resource an "invented archive," since it was drawn together for the specific purpose of creating an online resource [3]. Another example of an "invented archive" is the "CSS *Alabama* Digital Collection" at the University of Alabama, which uses a "virtual journey" image map to provide interactive access to documents and images associated with the Confederate raider, *Alabama*.

These "invented archives" take advantage of the connective capabilities of the World Wide Web and the links it can make among discrete texts, images, and other formats, bringing together materials even across repositories. At their best, the invented digital collections create a highly flexible, interactive, and engaging educational experience. Both the *Valley of the Shadow* and the *CSS Alabama Digital Collection* demonstrate that invented digital collections are powerful resources for teaching history. But the "invented archive" was not the model ultimately adopted for in-depth scholarly research.

Other repositories have created collection-based digital archives of primary materials that preserve the existing organizational structure of analog collections. These

primary source collections are captured in their entirety, and they have an important advantage. The intellectual integrity of an archival collection is based largely on provenance (meaning that records by one creator are not mingled with those by another) and maintaining original order. These two principles rest on the assumption that records are organized purposely by their creators and that the meaning of a document is derived from not only its content but its context within a group of documents. It supposes that there are relationships among records that should be preserved. The best online systems for managing and providing access to archival collections provide for both full-text or keyword searching and browsing of a hierarchy that preserves the original structure of the analog collection as described in the print finding aid.

The two largest examples of the collection-based digitization programs are the *American Memory* Project by the Library of Congress and the *Making of America* database by the University of Michigan and Cornell. *American Memory* currently provides access to 75 digital collections from the Library of Congress and other repositories. *Making of America* offers 1,600 books and 50,000 journal articles from the nineteenth century. But there are many other programs in place or emerging throughout North America. The Colorado Digitization Project is a collaborative initiative involving the state's archives, historical societies, libraries, and museums. The project includes statewide standards for applying metadata and regional scanning centers for image capture. The University of Southern Mississippi has initiated the "Civil Rights in Mississippi Digital Archive" project that will digitize and provide online access to manuscript collections, photographs, and other formats containing information on race relations in Mississippi. The first phase of the project is a digital collection of 125 oral history transcripts from civil rights figures such as Fanny Lou Hamer, Lawrence Guyot, and Hollis Watkins.

.04. Conclusion ([Return to Index](#))

The question of whether these types of digital collections can serve historians in "serious" research is partly dependent on the intended use of the resources. Is it the information a traditional resource contains or the object itself that actually constitutes the historical evidence? Digitization is sometimes insufficient for conveying physical attributes of a record, about the binding of a book or fine detail on oversize objects like maps. But as containers for content and basic physical attributes, digital surrogates are vastly more accessible and user friendly than microfilm, which remains the predominant reformatting option for primary materials.

Authority is also important. Establishing the reliability of primary sources is a fundamental part of the historical method. But just as in the analog world, historians will look to reputable information professionals to obtain the best online primary evidence on a given topic.

Also, the fact remains that effective electronic research only occurs if the evidence historians need exists online. Most primary sources will not be available on the Internet in the foreseeable future. The expense to digitize the entire holdings of a repository are simply too great. But while examination of physical records in libraries, archives, and other information repositories remains the central activity of historians, online research facilitates and sometimes supplements their traditional research. Large repositories in North America already provide significant online holdings. But it will be the degree to which the rank and file of informational repositories are able to provide online access to descriptive information about their collections in union catalog fashion and to digital surrogates of their most important holdings that will ultimately determine whether "serious" historical research can be done on the Internet in more than a limited way.

- [.05. References \(Return to Index\)](#)
- [.06. Selected Webliography on American History \(Return to Index\)](#)

P. Toby Graham, PhD
Head, Special Collections
McCain Library & Archives
The University of Southern Mississippi
email Toby.Graham@usm.edu
fax (601)266-6269
phone (601)266-5077
Box 5148
Hattiesburg, MS 39406-5148

© *The American Association for History and Computing All rights reserved.*

[Home](#) [Current](#) [Articles](#) [Works](#) [P-Resources](#) [E-Resources](#) [E-Journals](#) [Notices](#)