

Digital Libraries and Special Collections

Print for the People: A Mizzou Advantage Project

White-Paper 12/30/2010

Digital Libraries and Special Collections

Section 1: Introduction

Investigating ways for the University of Missouri to widen access to rare or special collections owned by the University is a primary goal of the Print for the People *Mizzou Advantage* project. An additional goal is to develop communities of scholars, curators, and other interested members of the university community to define and articulate the need for digital collections in research and teaching.

This report describes and analyzes a variety of digital collections and libraries created by other universities and government entities. Analyzing these collections can help create a stronger project model for the University and one that would suit the needs and resources of the University of Missouri. The evaluating of these existing open access collections are based upon the delivery of efficient and effective entrée for scholars, students, teachers and the public. This report will examine a variety of methods for digital document archives and provide a model for the next phase of the Print for the People project. The following analysis is split into three sections: local collections, national collections, and text embedded initiatives. Each digital collection is analyzed based on its content, size, and structure. Following the analysis is a summary and some recommendation for the next phase of Print for the People.

This paper is a work to which many have contributed. Alexa Wright and Zac Boesch, undergraduate students in the departments of English and the School of Journalism respectively were invaluable in researching, analyzing, and reporting on the many examples of digitized collections that exist in the digital world. The introduction and recommendations sections were contributed largely by Michael Holland, Director of the Special Collections, Archives and Rare Books Division (SCARaB) of the University Libraries. The Principle Investigators of this Mizzou Advantage grant project, Devoney Looser, Professor of English, Berkley Hudson, Associate Professor of Journalism, and Michael Holland as well as many members of the project steering committee contributed significantly to the content and the style of the document.

Section 2: Local Collections

Missouri Digital Heritage

<http://www.sos.mo.gov/mdh/>

Content: The Missouri Digital Heritage is a collection of digitized historical records pertaining to Missouri and the surrounding region. The Missouri Digital Heritage uses the Missouri Digital Heritage Initiative in order to digitize and make available online special documents housed in institutions around Missouri. The archive provides educational resources including lesson plans and annotated bibliographies to help teachers utilize those historical documents made available. The site also contains extended resources relating to but not included in the collections and exhibits made available already. Some of the larger archives include a searchable index of all death records from 1910-1950; the Soldier's database (contains images of original service cards)

a Virtually Missouri. The archives include maps, municipal records, government and political records, Missouri newspapers, photograph collections, and books and diaries.

Size: The archive has 324 collections and 38 exhibits digitized and available online.

Structure:

- **Organization:** The archive is divided into collections, exhibits and an education section.
- **Navigation:** The collections are most easily navigated by topic. Within each topic listing, the exhibits and collections are presented first by relevance (from high to moderate) and then in alphabetical order. For example selecting the topic of women presents one with collections of high relevance, or those which are primarily or only about the lives of women in Missouri, and then those of moderate relevance. For this reason, many of the collections and exhibits appear under several different topics. One can also view all of the collections and exhibits at once, organized by alphabetical order. When a particular collection or exhibit is selected, one is navigated to the site of the original content and the presentation of the documents varies depending on the source. Most include not only the original document but also a variety of background or pertinent information on the subject.
- **Searching:** The site has an advanced search function which allows a search throughout the website. The search can be narrowed down within groups like collection, date range, or county. Some of the material is not searchable including most of the contributed collections.

UM-Digital Library

<http://digital.library.umsystem.edu/>

Focus: The University of Missouri Digital Library offers text and image collections from UM libraries as well as hosting digital collections for other libraries and museums. The collections, both text and image, are centered on the history of Missouri as well as the history of the University of Missouri itself. Furthermore, the site contains a few links to other digital projects compiled and put on line by the University. The archive contains extensive information on the history of the University, including yearbooks, publications, and department histories. Many of the image collections feature local histories, such as postcards from the St. Louis World's Fair, and Plat books for every county in Missouri.

Size: The digital library is composed of 28 text collections and 35 image collections from the UM system and other institutions in the state.

Structure:

- **Organization:** The archived is organized into two separate categories: text and image.
- **Navigation:** The home page offers a variety of methods for browsing its contents. From the home page, the site maybe navigated by searching either primarily text or primarily image collections, or both together. The home page also offers two featured collections with an image sample and short descriptions of the material. The text and image collections are then listed individually so that the titles maybe browsed easily. Due to the small size of the collection this is the simplest way to search for a topic. When any particular item within a text collection is selected, one has the option of downloading the file, searching it, or viewing it page by page online. When a particular image within the collection is selected, the enlarged photo along with its description and historical reference point are presented.
- **Searching:** Collections themselves may be individually searched or searched in groups.

Washington University

<http://digital.wustl.edu/collections/>

Focus: The Washington University project focuses on materials in their libraries that are rare or difficult to access outside of the campus community. The focus of the Washington University Digital Gateway project is to create one interface for accessing the digital collections that the University already possesses. The Digital Gateway aims to encourage and provide a method for members of the University to further the digital collections by digitizing new material.

Size: The Digital Gateway contains just under 1700 texts in 10 separate collections along with links to additional resources for every collection.

Structure:

- **Organization:** The Digital Gateway is organized to allow both the searching and browsing of all the available collections through the use of drop down menus.
- **Navigation:** Selecting a collection navigates to separate page where the collection is housed. Some of the collections (for instance the LUNA architectural photos collection) require movement through 2 or more pages to access the collection itself.
- **Searching:** The search option allows the user to insert any word or concept and search through all of the digital collections for relevant materials. Each individual collection is structured differently, depending on the group. For instance, the LUNA collection can be searched or browsed in its entirety, while the Washington University Papyri collection can only be browsed through indexes by inventory or publication number.

Mercantile Library

<http://www.umsl.edu/mercantile/>

Focus: The St. Louis Mercantile Library was chartered by the state of Missouri and receives a grant from the National Endowment for the Humanities. The purpose of the grant is to enable the library to facilitate research and to create a modern organization of the archives. As a part of that advancement, the museum offers exhibits of its holdings that can be accessed online. The documents archived pertain to the cultural, political and aesthetic history of the American western frontier. The online exhibits represent primary image based holdings related to a particular theme. For instance, the “Adventures and Sufferings” exhibit presents historical images related to American Indian captivity narratives with the context of the story they depict.

Size: The museum currently has 9 online exhibits, both current and archived.

Structure:

- **Organization:** Each of the 9 exhibits within the online holdings are organized individually.
- **Navigation:** The exhibits maintain easy navigation through other organizational aides. For example, the “Sublime and Beautiful Mountain Scenery”: Western Images of A.E. Mathews” exhibit contains fifteen separate images which can be viewed and navigated in three separate ways (as a contact sheet, image by image, and by a map linking each image to the location it depicts.) Backtracking through individual collections can be laborious, as there is not the benefit of a search option for quickly moving between individual documents.
- **Searching:** The exhibits cannot be searched together, and some of the exhibits cannot be searched at all. However, considering the small size of the exhibits, this does not necessarily hinder the navigation or usefulness of the exhibits overall.

Truman Library

<http://www.trumanlibrary.org/>

Focus: The focus of the online holdings of the Harry S. Truman Library and Museum is to provide the user with important documents relating to some of the most critical historic events of the Truman presidency, including important international affairs such as the recognition of the state of Israel and domestic policy including the treatment of Japanese immigrants in the wake of WWII. The exhibits are an attempt to create complete and historically accurate depictions of some of the United States most important historical events.

Size: The size of the collection depends on the subject matter and could contain a wide variety of documents of varying prominence and relevancy. For instance the section on the friendship of Truman and Hoover contains not only pictures of the two presidents, a complete documentary history, oral histories and biographies, but also contains biographical sketches of people who worked under both presidents.

Structure:

- **Organization:** The online documents are organized by subject matter. Most of the online material is housed under the tabs “research” or “museum.” It is difficult to tell the difference between these two tabs and where things may be placed. Organization by subject matter makes scanning a variety of documents on a subject easier.
- **Navigation:** Selecting the subject of the “Recognition of Israel” leads the user to a page that is organized by more topics like “Seeking Zion” and “Jewish Homeland.” Each topic can be opened allowing for navigation to primary documents. The online holdings also include separate locations for the audio, political cartoon, and cinematic collections. For ease of use, seamlessly integrating these mediums into the collections by subject matter would make locating valuable research documents easier and more intuitive.
- **Searching:** Google sponsors the search but searching cannot be narrowed by collection or subject matter. This hinders research since searches get hits from the entire website.

Section 3: National Collections

Virginia Digital Library- Digital Collections

<http://virgobeta.lib.virginia.edu/catalog?portal=all>

Focus: The Digital Collections in the Virginia Digital Library cover a variety of topics from around the world. Some of collections are historical exhibits pertaining to the history of the South. Particularly notable is the Jackson Davis Collection of African American Photographs. The library also contains over 8,000 musical recording, though they can only be accessed by a login and password. The content of the digital library does not solely relate to regional or state history, however much of the content within particular collections relates to the history of Virginia and the beginning of the United States in general.

Size: The digital library contains 23 collections, with the size of individual collection ranging from one item in the SASKIA- SurveyImageSets collection to almost 12,910 in the Art and Architecture Collection.

Structure:

- **Organization:** The collections available are listed in alphabetical or in descending order of amount of material within the collection. The collections are presented without any description of what each contains.
- **Navigation:** The digital collections that may be accessed by the public are listed on the home page, though the larger online sources of the library seem to be only accessed from the University campus itself. When a particular article is selected, the full title, creator, date and collection it belongs to are presented and the full text or enlarged image may be viewed. All text documents may be viewed chapter-by-chapter or page-by-page, as some of them are rather large. They cannot be downloaded but do open in a separate window so that many may be viewed, if not searched, at once. When a particular article is selected a bar on the side contains links to related subjects.
- **Searching:** The entirety of the digital library can be searched by key word, author, title, or call number.

Documenting the American South (UNC)

<http://docsouth.unc.edu/index.html>

Focus: Documenting the American South, located at the University of North Carolina at Chapel Hill, offers digital primary resources on the history, culture, and literature of the American South. The goal is to provide rare and special documents for teachers, students and researchers. The project is an incredible resource on the American South during the Civil War, offering full textbooks on subjects as varied as social life in Virginia before the war to an antebellum slave narrative entitled *Twenty-Two Years a Slave, and Forty Years a Freeman*. Each text offers an extensive description of the narrative and the historical reference point it affords.

Size: The archives offer 14 collections with anywhere from a few hundred to more than 12,000 items.

Structure:

- **Organization:** The documents offered by DocSouth can be accessed by selecting the following: themed collection, alphabetical indexes by author, title or subject, or a general search of all of the sites holdings.
- **Navigation:** When the subject option is selected, it provides a detailed description of how to best navigate the subjects and is a very useful tool for finding a wide variety of information related to one topic over a longer period of time than any particular author or text might provide. Each individual entry consists of title, author, and content (for instance “1 printed document, 2 articles” or “3 images”), which gives the user easy access

to the size of the document they are opening before opening. Documents are offered in full text in either HTML or XML/TEI source files. Many of the HTML files of books and pamphlets show images of original covers and titles pages, though the text itself is reproduced type.

- Searching: Further materials not included in the themed collections can be searched by author, title or subject matter.

Georgia's Virtual Vault

<http://content.sos.state.ga.us/>

Focus: Georgia's Virtual Vault offers historical documents from the history of the state of Georgia from 1733 to the present. Many of the collections offer a survey of historical artifacts as the change over time, such as the Georgia Capitol Flag Collection or the pictorial reference available from comparing the Vanishing Georgia image collection with the Touring Georgia Collection.

Size: The vault offers 32 collections. Many of the collections contain large amounts of information, such as the collection of Georgia death certificates.

Structure:

- Organization: Collections are organized by subjects under the following headings: archives with images, without images, museum collections, and offsite collections.
- Navigation: The collections may be viewed in their entirety, or documents may be found within collections through the search or browse option. All of the collections are presented alphabetically by title on the left hand side of the page. When a particular collection is selected a short description appears followed by detailed search options. Individual search results may be selected and added to favorites at which point they may all be viewed simultaneously and retained as long as one remains connected to the site. Favorites may also be saved as a website, viewed in window with one another, or viewed as a slide show.
- Searching: The collections can easily be searched by keyword, time period, or descriptive characteristics. The browse option presents items by collection in alphabetical order complete with the title, an image, a record ID, and beginning date. One can also select the date option and have results presented according to their beginning date.

Making of America (Cornell University)

<http://digital.library.cornell.edu/m/moa/>

Focus: The Making of America Collection is a compilation of primary sources from the antebellum and reconstruction periods. The collection consists of digitized images of books and journals from this important period in American history. The archives host almost 20 different journals from the period dating back as far as an 1831 edition of New England Magazine. The materials provide a wide scope of information from many different ethnic and historical perspectives, though for the most part the documents are published texts, so for that reason perspectives can be somewhat limited by agendas.

Size: The number of documents available has increased vastly with the inclusion of journals, dating between 1837 and 1901, with anywhere from 40 to 200 volumes. The collection also includes over 267 monograph volumes.

Structure:

- **Organization:** The documents are divided into journals, monographs, and civil war documents that may be either searched or browsed. Each periodical in the collection of journals is presented first by volume and year, then by individual issue.
- **Navigation:** When a specific issue of journal is selected an interactive table of contents allows the user to select from front pages, editorials and advertisements. Pages are presented in the original text. Monographs are sorted by author. Some of the texts are over 1,000 pages long and can take a long time to load.
- **Searching:** Searching can be narrowed by collection or you can search the entire collection. Most documents allow for full text searching.

Early Americas Digital Archive

<http://www.mith2.umd.edu/eada/>

Focus: The Early American Digital Archive offers digital texts written in, or about, the Americas from as early as 1492 until 1820. The Archive is published by the Maryland Institute for Technology in the Humanities and seeks to integrate information from many different disciplines. The two main projects are the EADA Database and the “Gateway to Early American Authors on the WEB.” The archive contains documents from as early into American history as the journals of Christopher Columbus, to the poetry of Phillis Wheatley and Anne Bradstreet. The topics range from agriculture, Indian captivities, Calvinism, and slavery. Many of the later texts seem to focus primarily in the unfolding of events in the young United States, as opposed South American and Mexican history.

Size: There are 130 documents available in the archives. The number of documents available is smaller than many of the other archives due to their rarity. However some of the individual documents themselves may be over one hundred pages long.

Structure:

- **Organization:** The site is divided into the archive database and gateway features as well as a “News” section that outlines a list of newly offered texts complete with links to their locations within the database and a short description.
- **Navigation:** The texts themselves are plain text files, which make it possible to search with and across files for title, author, subject and specific terms. The two major research options are browsing alphabetically by author, which each letter linked to a separate page. This method is somewhat difficult unless you are looking for a specific text.
- **Searching:** The full text search function is extremely useful as it allows you select from a wide variety of criterion including person as subject, period, geographic location, genre, and format. One can also search for string of text throughout all of the texts in the archive.

Section 4: Text Embedded Initiatives

William Blake Archive (UNC at Chapel Hill)

<http://www.blakearchive.org/blake/>

Focus: The stated intended goal of the William Blake Archive is to effectively use the new medium of digital archiving to effectively represent the multimedia work of William Blake, where both the visual art and poetic of the artist can be displayed as intended. The archive declares a focus on the fundamental work and the poetry extracted from the totality of the work. The archive allows the work to be seen the way the artist intended, which, for most of the history of reprinting of Blake’s work, has not been the case as the medium required a focus on either the art, or the poetry. The site also contains a biography, glossary, chronology, and a scholarly article on illuminated printing.

Size: The archive contains many images and text documents of all of Blake’s 19 illuminated books, drawings, painting and engravings.

Structure:

- **Organization:** Blake’s works are organized by different categories like illuminated books or commercial book illustrations. The archives display reproductions of original works alongside thorough contextual information.
- **Navigation:** A tool called the Navigator simplifies navigating the site. The Navigator is a pop-out window that has all the categories of William Blake’s works. Once you click on a category a drop down menu appears of either Blake’s works or more categories that eventually direct you to the page the work is housed on.
- **Searching:** The archive offers both text and image searching. The text search is much easier to use and gives the user greater freedom, while the image search requires the user to select from amongst predetermined terms and categories.

Linda Hall Library

<http://www.lindahall.org/index.shtml>

Focus: The Linda Hall Library is an independent research library devoted to science, technology, and engineering. The content of the online collections and exhibit vary greatly. It includes an exhibit entitled The Face of the Moon: Galileo to Apollo and even digitized drawings of dinosaurs.

Size: The size of the collection varies based on the subject matter. The 13 digital collections contain different numbers of material and each collection could include numerous books or lengthy articles. For example, there are 128 images in the section of Dinosaur Discovery. The 10 online exhibitions also vary in size.

Structure:

- **Organization:** The online holdings are divided into digital exhibitions and digital collections. Digital Collections all have links leading to their separate page. The digital collections are organized by topic headings like Astronomy: Early Works or Philosophy of Science. The list of online exhibitions is not organized in a particular order.
- **Navigation:** The ease of navigation for a particular exhibition depends on the set up of the web page. While the Grandeur of Life is simple to navigate and very well designed, the Out of This World is difficult to navigate. Navigation of digital collections is fairly simple once a subject is selected. After selecting a book a table to the left allows for easy navigation to individual pages.

- **Searching:** Searching the digital collections allows for narrowing of search by specific collection. Most of the exhibits do not have a search feature but do have indexes.

Romantic Circles (University of Maryland)

<http://www.rc.umd.edu/>

Focus: Romantic Circles is an online study of the literature and culture of the Romantic Period. The content of Romantic Circles is made up of a variety of text documents including collected letters, novels, plays, travel journals and more.

Size: Romantic Circles is a large network of documents and online editions of text designed as a tool for studying the period in all its facets. The collection contains electronic editions, critical evaluations of texts, sound files and photos.

Structure:

- **Organization:** Romantic Circles is organized by: electronic edition, Praxis series, reviews, bibliographies, scholarly resources, and pedagogies. Most of the editions of texts are recreated text files, not screened images of the actual texts themselves and many of the text files have links to ‘footnote’ windows with outside information relating to particular ideas or concepts found in the document.
- **Navigation:** Navigation to each category is simple but under each category is a large list of all the documents in no particular order. After selecting a topic there is typically a table of contents.
- **Searching:** Searching is allowed for either the entire website, just a category, or just a single topic or edition.

Walt Whitman Archive (University of Nebraska- Lincoln)

<http://www.whitmanarchive.org/>

Focus: The Walt Whitman Archive is an intensive project designed to present educational and informational resources to create an online repository on the work and biography of Walt Whitman. Due to the existence of a large amount of diverse written texts, the archive contains notebooks, prose essays, letters, and journalistic articles. The site provides links to the archived materials, further resources, and a section for criticism. Many of the documents come from Libraries all around the United States, not just the University of Nebraska- Lincoln. The archive seems to lack the historical and social context that might provide more usefulness as a student

based research tool. Some of the fragments or sections seem under developed compared to others.

Size: The Whitman archive contains all seven editions of Leaves of Grass as well as German, Russian and Spanish Translations, two full-length biographies and an extensive collection of letters. The archive is smaller due to its narrowed focus.

Structure:

- **Organization:** The archived materials are divided amongst the published works and editions, manuscripts, biography and correspondence, criticism, and photo and sound files. Navigation:
- Navigation between pages of selected texts (especially scanned versions of original texts) can be extremely difficult, as each page of the document must be loaded as a separate web page. Published works, however, can be viewed either as complete e-texts, e-text of individual poems or page images. Each published work is provided alongside a critical essay on each particular edition and complete bibliography. The structure provides academic context and creates a thorough research tool.
- **Searching:** The search function is hosted by Google, and is very effective. Search terms bring up results that can then be refined by selecting the sector of the archive you want to search. For instance, one could search for the term ‘Emily Dickinson’ and then narrow the search by looking only through those documents filed under criticism.

Section 5: Summary and Analysis

The goal of Print for the People is to make materials more accessible to students and researchers. As demonstrated by many of the collections just digitizing material and putting it on the web does not necessarily make materials more accessible.

Local Collections

The local collections section provides insight into the various methods for presenting a focus on the history of Missouri and the larger Midwest. The large collections demonstrate what can be done with large collections of materials spanning extended time periods and large breadth of content. Missouri Digital Heritage (MDH) is an exceptionally well-structured digital archive, especially considering the amount and variety of information it contains. The MDH site manages to simultaneously function as a scholarly research tool and as a site for users with more general information needs.

National Collections

The national collections serve as an example of digitizing and organizing large quantities of information. The Virginia Digital Library, for instance, has an extensive digital collection of text and images; however, in the image collections such as the very large Holsinger Studio Photographic Collection, many of the objects are presented with little or no context. The date and location of the photo, as well as who or what is pictured, are displayed but there is no historical context further information presented. This problem often plagues large collections. One of the most interesting aspects of Georgia's Virtual Vault is that documents are accompanied by adequate historical context as the site attempts to digitally preserve the state's entire history in a balanced fashion.

Text Embedded Initiatives

Text embedded initiatives provide a very different model for archiving where the focus is on depth of content in a small subject area rather than breadth of subject matter and a variety of perspectives. The content choices are therefore shaped by the limited focus and become a matter of selecting criticism, historical context, or biographical information.

Electronic Scholarship

In order to gain the full advantage presented by the networking of local digitized collections, scholarship must establish its place in the digital academic world. Currently there are a variety of journals and monographs that are published in digital form and available from commercial and not-for profit associations. However, these resources are largely created in digital format and may or may not have a printed analog. There still largely using the format of the printed journals. The ability of scholars to publish digital scholarship has been an issue for many years. There have been some attempts to integrate digital scholarship with digitized resources and collections.

One of the most promising attempts to maintain scholarly quality and rigor is the NINES (*Networked Infrastructure for Nineteenth-Century Electronic Scholarship*) project. "Nines includes various kinds of content: traditional texts and documents — editions, bibliographic entries, and critical works of all kinds — as well as "born-digital" materials relating to all aspects of nineteenth-century culture. NINES is a model and working example for scholarship that takes advantage of digital resources and internet connectivity, while allowing scholars to integrate their contributions fully into their local IT environments. It provides scholars with access to a federated digital environment and a suite of computerized analytic and interpretive tools. A key goal of NINES is to go beyond presenting static images or transcriptions of manuscripts on-screen. Software tools that aid collation, comparative analysis, and enable pedagogical application of scholarly electronic resources expose the richness of the electronic medium."

http://www.nines.org/about/what_is.html

Section 6: Recommendations

- For digital humanities initiatives to move forward at MU the purpose and role of a digital humanities library must be clearly defined and articulated by scholars.
- The MU community that is interested in digital projects and collections needs to coalesce and continue periodic informational programs, such as the brown bag discussions sponsored by the Print for the People project.
- A stable and long-term electronic repository for digitized print materials needs to be established or located and designated.
- A model or several models for digitized collection management and access should be selected before further print materials are digitized and before a new website interface is created.
- Print and manuscript collections that are selected for digitizing at MU should be in the humanities and already physically owned by the university and/or its faculty members.
- Collections or materials selected for digitization must be in compliance with United States Copyright law.
- An organizational home for digitized collections needs be found on campus to replace the now defunct University of Missouri System (Library Systems Office) resources for digitization projects.
- A Center for Digital Humanities that provides infrastructure, technical support, and training for scholars desiring to engage in the creation of digital resources for research or teaching should be created and supported.
- A Center for Digital Humanities would also provide guidance to scholars interested in incorporated their scholarship and teaching resources into a federated digital environment such as NINES.
- Explore the practicalities of TEI (text encoded initiatives) at MU and attempt to achieve those standards used by the NINES coalition (Networked Infrastructure for Nineteenth-Century Electronic Scholarship) for collections of digitized materials.
- Navigation websites should remain intuitive in order to make research efficient and effective.
- Digital documents and objects must be easily searchable.