

Primary Research Material Online: Research and Teaching Perspectives

Paul J. O'Pecko

Mystic Seaport Museum, paul.opecko@mysticseaport.org

Abstract:

Digital imaging projects of the G.W. Blunt White Library at Mystic Seaport Museum, funded by the Library of Congress and the Andrew W. Mellon Foundation, are explained and demonstrated with comments on how the Library has incorporated the projects' outcomes into its workflow. Upon reapplying for additional funds, the Library was asked by the Andrew W. Mellon Foundation to instead reach out to scholars in the field and determine what their actual needs were in developing new scholarship. The Library hosted a conference and developed a survey to measure the scholars' needs. Methodology and preliminary results are discussed.

Keywords: Digital collections; humanities scholarship; primary sources; survey

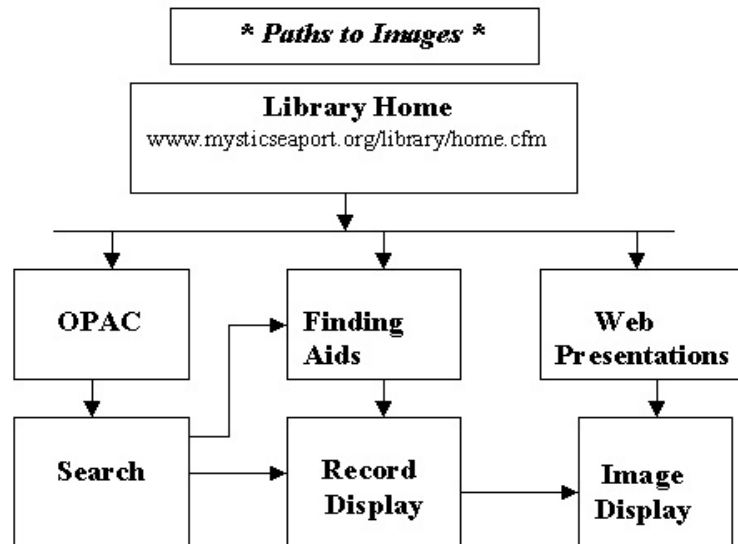
Introduction

The announcement last fall by Michael Keller, chair of the Digital Library Federation, that DLF is working on providing universal access to digital collections through way-finding and other means to facilitate use by scholars and others is the culmination of a shared thought process encompassing members and non-members of DLF alike. [1] Staff members in Mystic Seaport Museum's Library and Curatorial Departments have been working for more than a decade on integrating Museum and Library cataloging methods and records to provide an amalgamated, or federated, search capability for staff and outside users. The process has branched out into cooperative name and subject authority work, record management systems integration, digital imaging, information retrieval and the support of development of new scholarship through access to information. [2] The focus of this paper will be on the digital imaging projects and the survey whose results will help steer Mystic Seaport's efforts in contributing to both universal access to humanities resources and new scholarship.

Digital Projects

The G.W. Blunt White Library at Mystic Seaport has been involved in creating digital content for web display since 1998. The two major projects, one a Library of Congress *American Memory* project and the other an Andrew W. Mellon-funded project, total nearly 200,000 pages of images, mostly archival material and rare books. Library staff have attempted to incorporate digital work into daily routines and finding aids in addition to using the content as stand-alone "exhibits" and contributions to multi-institutional cooperative projects. This multi-versioning of digital assets is one way in which Mystic Seaport's Library effectively serves its various patron groups, including museum staff, graduate and undergraduate students, visiting scholars and the general public. Scholars and advanced students are adept at using the online catalog and EAD (Encoded Archival Description) finding aids, links from which take the user to the digital content, while at the same time the general public and others not familiar with the subject matter can explore the digital collections by browsing through a more

generic interface. The following schematic is a basic view of the different paths provided to the images in the digital collection.



Access via OPAC

A search of our online catalog, coincidentally and appropriately named Voyager from Endeavor Information, Inc., for the sample terms “Cunard” and “passenger?” will deliver a result set with a number of the bibliographic entries exhibiting an icon indicating the availability of an “electronic resource”. This electronic resource may consist of a single digital image, a collection of images, a collection of images with transcribed text, or possibly an associated EAD document, or finding aid for a collection of papers. Selecting the first record will take you to its complete bibliographic record in the catalog that includes such information as call number, author, title, subject headings and other typical entries. In addition to these typical entries is what you might consider an atypical entry, a field entitled “Linked Resources” with an active link to a digital version of the bibliographic entity. One of the functionalities that sets the site apart from a number of others is the ability for the user to easily page through the digital book or document or collection of documents once it is discovered that the item is of interest to the user. This functionality is enabled through the use of Microsoft Access and Coldfusion. All of the pointers to digital images as well as the images' related transcripts live in Access databases. We create active pages by using Coldfusion to query the databases, thereby creating an interface between html, the databases and the images. This process functions for all of the different digital formats, whether single-page documents, multi-chapter books or collections of papers separated into boxes and folders. Utilizing EAD documents to explore digital assets in collections can be understood by looking at one of the Library’s online collection registers.

Access via Finding Aids

A number of useful elements can be introduced using the guide to the *Fish Family Letters* (<http://www.mysticseaport.org/library/manuscripts/coll/coll178/coll178.cfm>).

This EAD finding aid is broken up into familiar segments including title, table of contents, overview, scope and content, etc. However, researchers using this collection get the added benefit of being able to link from the detailed description of the collection directly to images of the materials described. The goal in presenting materials in this way is to provide the user with an ease of operation that they would not experience in using the documents in person. It also achieves the added value of having the originals used less frequently and thereby better preserving them for future generations. In addition, we had the self-imposed requirement of digitally displaying the historical documents, giving researchers the feel for the original documents that a mere transcription would not provide. However, we also wanted to make it easier for those researchers to find what they were looking for by transcribing the documents into readable text and adding that text to the database thereby making the documents keyword searchable. Whether the researcher is looking at images of the original document or at the transcribed text, the ability exists to page through material as was seen earlier with the Cunard book.

Access via “Web Exhibits”

Presenting digital collections as stand-alone entities has been, and is, the typical method of display employed by institutions venturing into the electronic access world. Many organizations see this form of digital display as the climax of what they can achieve with online presentations. It is our belief that it is just one tool to be used to make users aware of the wealth of collections available for their use. Using this browse method for digital collections is by far the best way to introduce someone to subject matter that is foreign to them. Giving someone a search box or a card catalog utility will not help that patron find something that they do not know exists. If there is no frame of reference for someone coming blindly into a collection, they do not know what to look for. Providing a browse mechanism and additional context for the collection of images and text will go a long way towards making users comfortable with the subject area that will hopefully encourage further exploration. The Library of Congress' *American Memory* site facilitates this exploration by connecting similar digital collections through the use of a link entitled “related resources”. Clicking on this link will take the user to other digital collections featuring similar content.

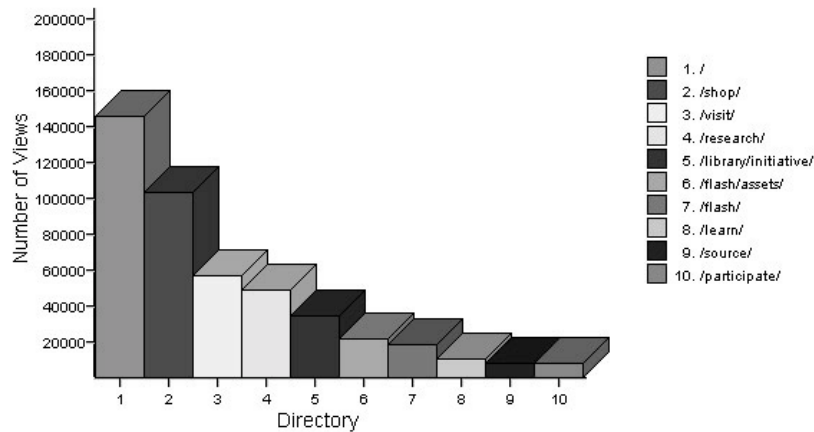
The first Mellon project helped the Library take the process used during the *American Memory* model and refine it a bit more. In addition to linking items to the catalogs and finding aids and adding transcripts, a great deal more content was added that would prove useful to novices, students and scholars alike. The site includes a broad overview of American maritime history and the significance of the digital collection in studying that history. Added to this are a number of hyperlinked bibliographies, ranging from a general bibliography of maritime history, to the most detailed bibliography in existence pertaining to voyages to the gold fields of California in the mid-nineteenth century. A timeline of American maritime history with links to appropriate digital content in the collections puts the collections in historical perspective. Items in the digital collection can then be accessed via a number of avenues, including subject access, title access, keyword searching, etc. One of the best stories involved with making this collection accessible has to do with non-paid volunteer data entry assistants. A group of over 30 volunteers from across the country were recruited to access pages of ship registers and index them for the project online. These registers span the second half of the nineteenth century and include information on hundreds of thousands of ships. This series is a

mainstay of the Library's reference collection. Volunteers added more than 1.5 million index entries to a database over a three-year period, enabling the user to search for a single vessel or, for the first time, its captain. If one searches on the term WARSAW, a result set will indicate the name of the ship, its captain, and the year and source of the information. Following the link will take the user to an image of the page from the actual register. Again, Coldfusion and Access databases provide the framework for searching and presenting the information and images. This site finishes out with a number of articles on maritime subjects that students and the general public frequently request. Drawn from the Museum's quarterly journal, these articles provide good, basic footnoted information for those interested.

Giving Users What They Want

As the following graph shows, the Library's digital initiative program (5) draws about 8% of the Museum's web traffic, or approximately 40,000 hits per month. The overall web traffic of the Museum is mainly geared towards the visitor (the Museum gets over 300,000 visitors per year) and the shopper at the Museum store.

Most Popular Directories



Based upon our success, we felt Mellon would provide further funding to continue the project. Mellon was unconvinced and suggested another course of action, namely asking our users what they wanted from us. We had initially built our case for a digitization project on the unique nature of our collections and the narrow use that they received because of the limited size of our institution. Many in the field of maritime history know of the collection and use it expertly, but the incredible expansion of electronic information gave us hope that we could reach many more of the uninitiated. The items we chose for inclusion in the project were selected using both objective and subjective means. We kept close track of what our users, especially our graduate and undergraduates used, noting both the specific items as well as the general subject matter, as we looked for trends. We identified heavily-used, deteriorating items that required scanning to save them from further decline through handling. But we also relied on instinct based on many years of working with the collections. We felt we understood best what people most frequently asked for or required. As stewards of the collections we felt the need to protect, but also to promote the collections. As Keller, et.al., observe, "By serving as custodians of local collections, they [librarians] have

incidentally served a larger common good. Whatever other public benefits they provide, publishers and Internet promoters do not, and cannot be expected to, fulfill this custodianship role.” [3] We were attempting to be not only the custodians but the publishers and providers as well without seeking the proper assistance in decision-making during the item selection and display selection phases. What this all came down to, as Mellon rightly concluded, was that all of our selection criteria and methods of access were self-selected. We were giving our users what we felt would benefit them the most rather than asking them what would be most beneficial to their research. In an article on e-resource trends, Okerson frankly says, “Our profession should do what our commercial information suppliers are doing: focus on the users, their needs, their wants, and the practices of using information.” [4] This, in essence, was Mellon’s challenge to us regarding the maritime studies field and, in particularly, scholars’ use of primary resources online in pursuit of developing new scholarship.

We took on Mellon’s challenge and decided to try and work more closely with academics in the fields of maritime studies, especially those involved in history, archeology and literature. With funding from Mellon we hosted a meeting of approximately thirty scholars from universities around the U.S. and Canada (see <http://www.mysticseaport.org/library/about/edrindex.cfm>) in September 2003 to begin the discussion on exploring online digital resources for use in research and teaching. The spirited discussions that ensued on these topics over a two-day period provided the background information we needed to create specific questions for a survey. An added benefit of the meetings was an agreement that Mystic Seaport would begin holding biennial conferences in maritime studies beginning in 2005. Having a recurrent conference provides a frequent opportunity for outreach sessions for those most likely to use our resources. Tibbo stresses merely making materials available online is not enough. Outreach and education are critical in making “access accessible.” [5]

Surveying Academic Patrons

In creating a survey that built upon the conference, but also suited the Library’s needs, a number of previous surveys were reviewed to take advantage of their findings. Reports by Amy Friedlander [6], Carol Tenopir [7] and Helen Tibbo [8] were extremely useful in providing a baseline of knowledge where use of electronic resources was concerned. Data on populations of users, amount of time making use of sources, knowledge of sources, retrieval strategies in various environments, etc. is presented in great detail. We hoped that any survey we produced would have practical and immediate impact on our actions and current and future digital projects. Unlike university programs in Library Science that develop surveys as teaching tools, as theoretical models as well as for practical applications, we felt the need to concentrate on the latter in order to better serve our core constituencies through the creation of products they had a hand in choosing. Their information needs and desires will drive our digital programs forward.

The Survey (<http://www.mysticseaport.org/library/survey/survey.cfm>)

The first section of the survey was mainly posed in order to guarantee that our survey audience was what we expected it to be. In addition to sending the link to the survey to specific individuals, we posted it on humanities listservs, opening it up to a wider group of participants. We attempted to corroborate earlier findings that scholars would indeed make use of digital sources whether the original primary resources were five

hundred miles away or fifty. One of Tenopir's clear conclusions is, "Both faculty and students use and like electronic resources and most readily adopt them if the sources are perceived as convenient, relevant, and time saving to their natural workflow." [9] 65% of our initial respondents were strongly in this camp.

Our discussions with scholars during the conference led us to concentrate our efforts on discovering which avenue of pursuit in electronic resources they would truly prefer us to follow in supporting both research and teaching. It was made very clear that there were three specific areas of interest in their search for quality electronic research information: Resource Discovery Tools, Demographic Databases and Digital Historical Newspapers. With the groundwork laid from studies such as Tenopir's, and expected validation from ours, that scholars were at ease with online sources, the next step in our process was to determine which type of Resource Discovery Tool would be most used, if not necessarily the most useful, to them in their search. One direct and one indirect set of questions were added to the survey. The first set of direct questions dealt with the nature of their personal searching preferences. If given the chance to choose, would respondents prefer a free text search mechanism akin to *Google*, a subject search similar to an online catalog, or an Index/Directory model that could be compared to a more-refined *Yahoo*? Initial response to this straightforward set of questions indicates the *Google*-type free-text search as the favorite, garnering 55% of the top votes. We then hoped to determine whether scholars were interested in a comprehensive discovery of information as opposed to what can generally be found on the web. Asked whether they would be more interested in learning the details of a 600 item collection directly related to their field of study, or viewing a set of 100 of those items online (and not being informed that there were 500 other items), 78 % chose the comprehensive listing over the convenience of the online images. However, the predominant follow-up comment to this question was that they would obviously prefer having all 600 items available digitally, adding credence to the earlier response that 65% were willing to use digital images in place of the original materials when available.

Demographic databases also stimulated prolonged discussion at our September 2003 meeting. Historians, especially, find databases of information such as census results useful, particularly if the results can be sorted according to the scholar's particular needs. There are various demographic records kept in maritime studies, including lists of ships' crews over a wide range of time. With this information in hand we hoped to establish what fields of information would be most useful and whether scholars would be more interested in a thorough indexing of a smaller subset (eg., one year's census), or access to images of a larger subset (eg., census data from four separate years) with only limited indexing. Scores for the particular fielded information scholars wished to sort on were predictable to a large degree and, as with earlier results in Resource Discovery, scholars preferred the comprehensive nature of indexing for the single-year database, rather than the broader appeal, though lesser accessibility, of the four-year grouping.

The third area of interest identified by our core group of scholars was that of digital newspapers. Again, based on discussions, we were working on the given that scholars wanted newspapers available digitally. What we were attempting to determine was the delivery and search formats in which they were most interested in order to guide our digital projects. Keyword searching capability once again led the field followed, in

order, by transcribed text, full images of the newspaper, date search and sort, and lastly searching by subject indexing. In a final question on priority ranking we asked our users to tell us which topic in our survey was most important to them. Not surprisingly, initial results place Resource Discovery Tools at the top of the list followed by Digital Newspapers and Demographic Databases. Knowing that a source exists, and where to find it, is more desirable than drilling down into that source electronically.

Classroom Tools

As we learned at our conference, more and more professors are making use of electronic resources in the classroom. And although streaming video and audio are pervasive in today's electronic environment, the survey results showed that when presented with such alternate digital teaching tools, most academics in the classroom will continue to rely on what is familiar to them in terms of cultural heritage materials. Topping the list of teaching aids are images of manuscript material, historic photographs and images of subject-related books. Other electronic formats, whether maps, audio, video, etc. are currently lower on the scale.

Conclusions

Aside from helping to guide us in preparing specific digital products for use by scholars in our field of interest, the preliminary results of this survey verified some of our beliefs and backed the findings of earlier surveys. First, comprehensiveness in searching in one's area of study outpaces other needs. Secondly, ease of use (eg., keyword searching) is mandatory in any system that is developed. And finally, tried and true practices of librarians and archivists, as Tibbo advises [10], often run counter to 1 and 2. These practices may be seen as irrelevant to the general population of scholars seeking appropriate information. The creation of subject headings, indices, catalog records, EAD documents, etc. can serve a purpose, if they can be initially accessed via keyword searching on the web, but are otherwise overlooked by researchers going online to track down information.

- 1) Digital Library Federation Fall Forum, Albuquerque, New Mexico 17-19 November 2003: A Summary.
URL: <http://www.diglib.org/forums/fall2003/ForumNov03summ.htm>
- 2) Prescott, Leah. 2001. Mystic Seaport in the Information Age.
URL: <http://www.mysticseaport.org/library/msitia/intro.html>
- 3) Keller, Michael A., et al. 2003. What is a Library Anymore, Anyway? First Monday , 8.
URL: http://firstmonday.org/issues/issue8_5/keller/index.html
- 4) Okerson, Ann. 2000. Are We There Yet? Online E-Resources Ten Years After. Library Trends 48 (Spring,): 671-693.
- 5) Tibbo, Helen. 2003. How Historians Locate Primary Materials: Educating and Serving the Next Generation of Scholars. ACRL Annual Conference, April 10-13, 2003. Charlotte, North Carolina.
URL: <http://www.ala.org/ala/acrl/acrlvents/tibbo.PDF>
- 6) Friedlander, Amy. Dimensions and Use of the Scholarly Information Environment: Introduction to a Data Set Assembled by the Digital Library Federation and Outsell, Inc.

Washington, D.C.: Digital Library Federation and Council on Library and Information Resources, 2002.

URL: <http://www.clir.org/pubs/reports/pub110/contents.html>

- 7) Tenopir, Carol. Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies. Washington, D.C.: Council on Library and Information Resources, 2003.

URL: <http://www.clir.org/pubs/reports/pub120/pub120.pdf>

- 8) Tibbo, Helen, et al. Primarily History in America: How US Historians Search for Primary Materials at the Dawn of the Digital Age. Pre-publication draft. University of North Carolina at Chapel Hill. 2003.

- 9) Tenopir, Carol.

- 10) Tibbo, Helen. 2003. How Historians Locate Primary Materials