



DLF SCHOLARS' PANEL

Wednesday June 2-Thursday June 3, 2004
Washington, DC

AGENDA

The impulse driving this meeting was in a large part a desire to learn from working scholars what they value and what they need from our digital library services -- to test and temper our sense of what library users want against the expressed needs of scholars. Everyone attending is deeply engaged in creating and using digital library content, often in partnership with an academic library.

ATTENDEES

- Vernon Burton. Professor of History and Sociology, and Senior Research Scientist, National Center for Supercomputing Applications, University of Illinois at Urbana Champaign.
http://www.history.uiuc.edu/fac_dir/burton_dir/burton.htm
<http://riverweb.cet.uiuc.edu/>
- Morris Eaves. Professor of English and Project Director of *The William Blake Archive*, University of Rochester.
<http://www.rochester.edu/college/eng/faculty/eaves.html>
<http://www.blakearchive.org>
- Mark Kornbluh. Professor of History and Director of MATRIX: Center for the Humane Arts, Letters, and Social Sciences OnLine, Michigan State University.
<http://www.history.msu.edu/faculty/kornbluh.html>
- Jeff Looney. Editor-in-Chief, Papers of Thomas Jefferson, Retirement Series. Thomas Jefferson Foundation, Charlottesville, Virginia.
<http://www.monticello.org/papers/index.html>
- Thomas Luxon, Associate Professor of English and Director, Dartmouth Center for the Advancement of Learning, Dartmouth College.
<http://www.dartmouth.edu/~english/faculty/luxon.html>

- Kenneth M. Price, Professor of American Literature and co-editor, *The Walt Whitman Archive*, University of Nebraska, Lincoln.
<http://www.unl.edu/Price/>
<http://www.whitmanarchive.org/>
- Stephen Railton, Professor of English and Director of *Mark Twain In His Times: An Electronic Archive*, and *Uncle Tom's Cabin & American Culture: A Multi-Media Archive*, University of Virginia.
<http://etext.lib.virginia.edu/railton/railtonhp.html>
<http://etext.lib.virginia.edu/railton>
<http://jefferson.village.virginia.edu/utc>
- Roy Rosenzweig, Professor of History & New Media and the Director of the Center for History & New Media, George Mason University.
http://cas.gmu.edu/historyarthistory/faculty_staff/biography.php?f=4667
<http://chnm.gmu.edu/>
- Benjamin C. Ray, Professor of Religious Studies and Project Director, *Salem Witch Trials Documentary Archive*, University of Virginia.
<http://www.salemwitchtrials.org/>
- Crandall A. Shifflett, Professor of History and Director of Graduate Studies at Virginia Polytechnic Institute and State University; Project Director, *Virtual Jamestown*.
<http://www.virtualjamestown.org/>
- Martha Nell Smith, Professor of English and Director of MITH, University of Maryland; and General Editor, the *Emily Dickinson Electronic Archives*
<http://www.mith.umd.edu/mnsmith/>
<http://www.emilydickinson.org/>
- William G. Thomas, Associate Professor of History and Director of the Virginia Center for Digital History, University of Virginia.
<http://www.virginia.edu/history/faculty/wthomas.html>

In addition, the following digital library specialists were in attendance:

- Nancy Davenport (Library of Congress)
- Barrie Howard (DLF)
- David Greenbaum (UC Berkeley) Co-chair
- Leslie Johnston (U Virginia)
- David Seaman (DLF) Co-chair

FINDINGS

Introduction

College and university librarians have a long tradition of listening to their users, and of adjusting our services and collections according to the articulated needs of our faculty and students. This is nowhere more important than in our emerging digital library endeavors, where much is still unfamiliar to patrons and where new products, aggregations, tools, and services come (and often go) with confounding frequency. Across the many initiatives, benchmarks, and standards undertaken by the Digital Library Federation (DLF) is an overarching desire to build library services and online holdings that result in richer scholarship and more effective pedagogy.

To help inform this end, and to test our own assumptions about developing needs in digital scholarship, the DLF convened in the summer of 2004 in Washington, DC, a group of humanities and social science practitioners, all of whom are actively building digital archives, online editions, and electronic scholarship to further their academic and teaching interests, and who are working with their library colleagues and digital collections in innovative ways. Over two days of lively and free-flowing discussion in June, the scholars provided feedback on how libraries could partner with them to serve their particular digital scholarship needs.

Follow-up discussions with several participants have further fleshed out the themes and observations detailed below.

Barriers to Digital Scholarship

We turned our attention early to the hurdles that face this first wave of scholars undertaking serious digital scholarship, in order to understand which of these barriers can be overcome by emerging digital library research. There was speedy and widespread agreement that an overarching problem was the lack of persistent identifiers -- permanent and trusted internet addresses -- for online objects. How can you invest in rich, hyperlinked scholarly writing or scholar-driven archives if the material not under your immediate control keeps moving from web address to web address, or disappearing altogether (an irritation commonly known as "link rot")? It is a waste of time to have to monitor and fix broken links, and a disincentive to undertaking further work. This is a problem they look to libraries and publishers to solve, and to solve quickly ("aren't you guys supposed to be good at this sort of standardization" one participant said). As a positive example of persistent identification in the scholarly journals industry we looked at Crossref and the Crossref/Google article search service, which have grown up around the Digital Object Identifier (DOI) persistent ID that is commonly used in the STM scholarly journal arena.

The other main thread of this conversation about hindrances to digital scholarship -- which came up in discussions of institutional repositories too -- was the failure of departmental promotion and rewards structures to recognize and accommodate the shift

from a print-based to a digital world of scholarly publishing and communications. It is no accident that most active humanists and social scientists working with digital media are post-tenure, one participant observed, and I suspect that even then they are not all immune from the career-depressing effects of being seen to be "too digital" or "only digital."

Need for Tools

The group was clear that there is a severe need for tools customized for a range of scholarly inquiry needs:

- Gathering information from multiple sources, along with some information about it (personal libraries with metadata)
- Searching of images
- Visualization of patterns and trends and search results
- Annotation of text, image, and multimedia files
- Writing the new scholarship -- authoring tools for the digital scholar

However, so unfamiliar is this area that we heard from several individuals that they had a hard time articulating precisely what they required from such tools, or what level of software creation skills or consultancy is available to them, and where. We are still in a stage where it is easier to react to an example of an existing tool than to dream them up ex nihilo, and with that in mind we discussed and demonstrated a variety of software packages that allowed scholars to gather, search, annotate, and re-package digital objects from library collections, including New Zealand's impressive Greenstone (referred to in this context as a personal library organizer), the suite of tools from UC Berkeley's Scholar's Box initiative, and Michigan State's Matrix annotation software that is aimed at various streaming media. Clearly a first-order need for this group was simply to know how to discover that these sorts of products exist (let alone the range of locally created but re-usable software custom-built for various initiatives), and what their characteristics are.

Services: Repositories and Harvestable Metadata

There has been a rapid growth in the ambitions of universities to build systems to safeguard and re-use the full range of scholarly and pedagogical output -- the institutional repository movement. Opinions about this phenomena may well differ across disciplines; for this group there was a decidedly cool reception to the notion of turning over their scholarship, datasets, and archives to their institution for exploitation as institutional assets (the language of the institutional repository discussions may well be to its detriment -- faculty do not necessarily take kindly to being cast as asset workers producing exploitable product for their institutions, even if only at the level of language). While the ability to have a long-term safe-haven for their digital content found some real favor, especially as it was curated by the library, there was a range of concerns beyond this -- questions of ownership, permissions, load (how much work is it to prepare a body

of material for a repository?) and again the observation that there was no link between the re-use of a scholarly asset and the current faculty rewards system.

Much more positive was the reaction to sharable and harvestable metadata -- not a concept that was very clear to the group prior to the meeting. We used the Open Archives Initiative (OAI) as an example of simple metadata records for digital objects that are put on the web and harvested by software, in order to build services that include records from many sites all arranged in one service of portal. There was a good deal of interest in this mechanism both as a way to help make their own scholarship more visible, and as a way of gathering up references to related material to which they may want to refer.

Digital Library Collections

Given the active involvement these scholars have in building and contextualizing content -- in engaging actively in the creation of digital archives that they then manipulate -- and given the concern with link rot -- it was no surprise to learn that a behavior they wanted from collections of digital objects was the ability to capture and re-use that material in their own local contexts. There was firm agreement that it is not always enough to link to a resource in someone else's system, even if the link is persistent. The need for a local copy may be aesthetic integration into an archive; offline use; incorporation into a desktop tool of some sort (data visualizer; annotation tool; courseware package; textual analysis software); data enrichment with terminology of the scholar's choosing; or even the simple need to search a body of material all at once -- impossible when the books are in different systems with different search tools. Equally clear is how difficult it is to get permission from data holders to satisfy this common need, even when the material in question is freely available on the internet in archives and libraries. Typically the institutions who digitized and who host the material do not have policies in place, or rights expressions, to allow that content to have a secondary life in an online project at another institution. "Just link to it" is not the answer often for this group of scholars, but absent a mechanism to explicitly accommodate the desire to bring digital objects into a local scholar's archive, they are left with a frustrating and time-consuming series of conversations, favors, and personal pleas in order to engage deeply and actively with the material in digital library collections.

Closing

Work with this group has been lively and enlightening -- for individual projects and in an ad hoc manner for the organization as a whole. Such scholarly users make for very effective reaction and review panels. After the event, several members articulated a need for help in acquiring either digital copies of items as yet undigitized or the permission to move digital items held elsewhere into their own archives and tools. The latter may well give us a clearer sense of how and when simple access is not enough, and close engagement with and enrichment of a file in another library's collection is what is needed to fulfill a scholarly or pedagogic need. In addition, one specific opportunity for partnership was put forward, by the Virtual Jamestown group; I enclose this as an Appendix in case it touches a nerve with any DLF library.

BIBLIOGRAPHY

The following topics, complete with web addresses and annotations, are culled from -- and act as references to -- our free-ranging conversations during the June 2004 DLF Scholars' Panel meeting.

- [1.](#) Federated searching between Google and Amazon, Google and university repositories, and Google and OCLC
- [2.](#) Mass Digitizing Efforts and Ambitions
- [3.](#) Institutional Repositories and Digital Library Repositories
- [4.](#) Digital Library and Scholarly tools
- [5.](#) Data Sharing (datasets)
- [6.](#) Shareable Metadata
- [7.](#) Courseware
- [8.](#) Persistent Identifiers
- [9.](#) Scholarly Publishers and Archiving
- [10.](#) Digital Preservation
- [11.](#) Online Communities
- [12.](#) Legislation
- [13.](#) Articles that Exemplify Types of "New Scholarship".

1) Federated searching between Google and Amazon, Google and university repositories, and Google and OCLC

A9

<http://www.a9.com/>

Marriage of Amazon and Google which is analogous of any single search that links online and print materials.

DSpace

<http://chronicle.com/free/2004/04/2004040901n.htm>

Read an article about how this consortium plans to use Google to search the contents of institutional repositories.

OCLC

<http://www.oclc.org>; <http://www.oclc.org/worldcat/pilot/>

Read about their experiment to add 2 million WorldCat records into Yahoo! And Google with zip code lookup of the nearest place they know about that holds the book.

2) Mass Digitizing Efforts and Ambitions

Digital Promise / Digital Opportunity Investment Trust (DO IT)

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

A proposed \$20 billion trust fund from the sale of Federal airwaves, to be held by the U.S. Treasury, the interest from which (circa \$1 billion per annum) would drive massive digitizing for the public good, tools building for teaching and using digital content, assessment of digital learning, etc.

Government Printing Office (GPO)

http://www.gpoaccess.gov/about/speeches/04232004_IST.pdf

This agency plans to digitize all non-digital Government Documents and make them electronically available.

3) Institutional Repositories and Digital Library Repositories

DSpace Federation

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

The best-known of the "institutional repositories" -- an open source digital library system to capture, store, index, preserve, and redistribute the intellectual output of a university's research faculty in digital formats, developed jointly by MIT Libraries and Hewlett-Packard.

Fedora Project

<http://www.fedora.info/>

An Open-Source Digital Repository Management System, used currently to manage local digital library content, and with growing ambitions to operate as the underpinning for institutional repositories and content management systems. Developed by Cornell University and the University of Virginia.

4) Digital Library and Scholarly tools

Greenstone Digital Library Software

<http://www.greenstone.org/cgi-bin/library>

A free "digital-library-in-a-box" delivery program, written in New Zealand specifically for use by libraries and scholars online, on a PC, or on a CD. Ongoing development -- METS and OAI abilities being added (see Shareable Metadata, below, for a sense of OAI and METS), as well as richer XML import capabilities. To see an example of Greenstone being used on-line in a DLF library, see *Chopin Early Editions*: "A collection of digital

images of early printed editions of musical compositions by Frédéric Chopin. This collection was created by the University of Chicago Library and, once completed, will include its entire collection of over 400 Chopin early editions."

<http://chopin.lib.uchicago.edu/>

Scholar's Box [David Greenbaum, UC Berkeley Interactive University]

<http://iu.berkeley.edu/IU/SB/>

Demonstrated at the meeting.

Media Matrix [Mark Kornbluh, Michigan State University]

<http://www.matrix.msu.edu/digital.php>

Demonstrated at the meeting.

Center for History and New Media tools [Roy Rosenzweig, George Mason University]

<http://chnm.gmu.edu/tools>

Discussed at the meeting. More recently, Roy has announced the **Echo Tools Center**: "The number of historians interested in using digital tools to facilitate their work has been rapidly expanding, as has the number of researchers developing online tools for the humanities. In order to facilitate contact between these two groups, Echo would like to announce the beta launch of its new Tools Center, an experimental, comprehensive resource for scholars interested in the nuts and bolts of online history. Just as Echo's Research Center offers a guide to thousands of history websites, the Tools Center is envisioned as a central directory of the myriad pieces of software and other tools available to contemporary historians. Built using the same open-source software that powers sites like Wikipedia, the Tools Center is a specifically collaborative resource, enabling developers to post descriptions of their products, and users to apply their own expertise to build and expand its entries. Though still in beta form, we invite both historians and software developers to visit the Tools Center at <http://echo.gmu.edu/toolcenter-wiki/> and contribute their knowledge to this growing asset to the online history community."

5) Data Sharing (datasets)

Inter-University Consortium for Political and Social Research (ICPSR)

<http://www.icpsr.umich.edu/access/deposit/guidelines.html>

For 40 years, an increasingly vast archive of social science data for research and instruction. "ICPSR preserves data, migrating them to new storage media as changes in technology warrant. In addition, ICPSR provides user support to assist researchers in identifying relevant data for analysis and in conducting their research projects."

6) Shareable Metadata

Open Archives Initiative (OAI)

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

This allows one to create a metadata record for an object, and "publish" that catalog record in a space on a web server that allows others to "harvest" it; having gathered up lots of records from lots of locations, one can use them to create a service -- a one-stop-shop for a topic or discipline, for example (see *NSDL* below). Typically, the records contain information about the object's date, author, title, genre, subject, etc. potentially allowing a rich service to be created (although the service is only as good as the metadata it aggregates, which can vary wildly in accuracy and completeness from site to site, complicating the creation of a robust and trusted service).

National Science Digital Library (NSDL)

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

This is a subject-focused OAI service -- "a digital library of exemplary resource collections and services, organized in support of science education at all levels." The content resides on many different sites, but OAI records are harvested from them to a central location to build the NSDL "portal".

OAIster [University of Michigan Digital Library Production Services]

<http://www.OAIster.org>

A collection of all OAI records that have online publicly accessible content. Searchable by author, date, subject, etc.

7) Courseware

SAKAI

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

The Sakai Project is a \$6.8M software development project founded by The University of Michigan, Indiana University, MIT, Stanford, the uPortal Consortium, and the Open Knowledge Initiative (OKI) with the support of the Andrew W. Mellon Foundation. The project is producing open source Collaboration and Learning Environment (CLE) software.

Open Courseware Initiative

<http://ocw.mit.edu/index.html>

700 online university courses from MIT, which are free on the Web. For a full list of courses, visit <http://ocw.mit.edu/OcwWeb/Global/all-courses.htm> .

8) Persistent Identifiers

Note: These variously work by having the user point not to a file on a website but to a

unique ID number in a web-based service that matches the unique ID with a web address. As long as you update the central registry, you can move files from web address to web address without any links breaking. Enter the following URL into the address bar of your Web browser -- or press Ctrl + click with your pointer over the UR -- to see how this works, <http://dx.doi.org/10.1037/0003-066X.59.1.29>.

PURL

www.purl.org

This service is free and used by some libraries. It works only for web-based information.

Digital Object Identifier (DOI)

<http://www.doi.org>

Used by many scholarly publishers, DOIs can be assigned to anything. It costs money to use.

See also DOI FAQs at

<http://www.tsoid.co.uk/FAQ.aspx?PageID=FAQ&category=general> ;

DOI and Google at <http://www.crossref.org/01company/pr/press20040428.html>

Radio frequency identification (RFID) tags

<http://news.com.com/2010-1069-980325.html>

<http://www.rfidupdate.com/>

<http://www.rfidjournal.com/>

9) Scholarly Publishers and Archiving

Elsevier

<http://www.elsevier.com/> ; www.icsti.org/forum/46/Joep_Verheggen.ppt

Read about their relationship with the Dutch Royal Library (Koninklijke Bibliotheek).

Elsevier now allows all its authors to publish their Elsevier-published articles for free on the web. Read the article about Elsevier's open access self-archiving at:

<http://www.infotoday.com/newsbreaks/nb040607-2.shtml>

Blackwell

See also "Blackwell Publishing Ltd and the Koninklijke Bibliotheek Sign Archiving Agreement," <http://www.blackwellpublishing.com/press/pressitem.asp?ref=83&site=1>

10) Digital Preservation

National Digital Information Infrastructure for Preservation (NDIIPP) [US]

<http://www.digitalpreservation.gov/>

A federally-funded project, administered by the Library of Congress, to explore and design a distributed infrastructure for preserving "born digital" materials.

See also:

Digital Curation Centre (DCC) [UK]

<http://www.dcc.ac.uk/>

Digital Preservation Coalition (DPC) [UK]

<http://www.dpconline.org/graphics/>

Preserving Access to Digital Information (PADI) [Australia]

<http://www.nla.gov.au/padi/>

11) Online Communities

H-Net

<http://www.h-net.msu.edu/>

Mark Kornbluh and colleagues, at The Center for Humane Arts, Letters, and Social Sciences Online, Michigan State University: 100 free electronic, edited, interactive newsletters; over 100,000 subscribers in more than 90 countries.

12) Legislation

Public Access to Science Act (HR 2613) -- The Sabo Bill

SPONSOR: Representative Martin Olav Sabo D-MN

STATUS: Introduced in the House, 06/26/03

Amends Federal copyright law to declare copyright protection unavailable to any work produced pursuant to scientific research substantially funded by the Federal Government.

Requires any Federal department or agency that enters into a funding agreement with any person for the performance of scientific research to include in the agreement a statement that copyright protection is not available for any work produced pursuant to such research under the agreement. Expresses the sense of Congress that any Federal department or agency that enters into such funding agreements should make every effort to develop and support mechanisms for making the published results of the research conducted pursuant to the agreements freely and easily available to the scientific community, the private sector, physicians, and the public.

13) Articles that Exemplify Types of "New Scholarship".

Thomas III, William G., Edward L. Ayers. "The Difference Slavery Made: A Close

Analysis of Two American Communities." University of Virginia.

<http://www.vcdh.virginia.edu/AHR/> (accessed June 10, 2004).