

The Distributed Library: OAI for Digital Library Aggregation

OAI Scholars Advisory Panel Meeting

June 20-21, 2005

Washington DC

Note: while this Panel includes some members of the June 2004 DLF Scholars' Panel (<http://www.diglib.org/use/scholars0406/>), the two meetings had quite different purposes; the former was a wide-ranging discussion of a variety of issues; this one is much more focused on the needs, opportunities, and obligations of the DLF IMLS grant.

The current IMLS grant to the DLF—*The Distributed Library: OAI for Digital Library Aggregation*—has as an important part of it a pair of advisory committees, one technical and one scholarly. This meeting was the first face-to-face gathering of the OAI Scholars' Advisory Panel.

Composition of the Panel: The Scholars' Panel is comprised of teaching faculty from DLF members and other major institutions.

Purpose of the Panel: These scholars are well-positioned both to speak to scholarly user needs and interests across all levels of university constituencies (i.e., undergraduate, graduate, faculty, and staff), and to comment from an end-user perspective on the prototypes and experimental services we develop. The Scholars' Panel will serve not only as a key resource for evaluation but will also contribute substantial subject expertise during the content selection/collection development phase of the project.

OAI Scholars' Panel Members

Gail McMillan, Virginia Tech., Director, Digital Library and Archives
<http://scholar.lib.vt.edu/staff/gailmac/Gailshp.html>

Ken Price, U Nebraska, Lincoln, English: Walt Whitman
<http://www.unl.edu/Price/>

Roy Rosensweig, GMU History and Center for History and New Media: scholars tools
<http://chnm.gmu.edu/>

Bruce Rosenstock, UIUC, Associate Professor of Religious Studies and the Coordinator of Computer-Assisted Instruction of the College of Liberal Arts and Sciences.

Steve Railton, UVa, English: Harriet Beecher Stowe; Mark Twain
<http://etext.lib.virginia.edu/railton/railtonhp.html>

Martha Nell Smith, U Maryland, English and MITH: Emily Dickinson
<http://www.mith.umd.edu/mnsmith/>

Randy Shifflett, Virginia Tech., History: Virtual Jamestown
<http://www.virtualjamestown.org/>

Will Thomas, UVa, History: Civil War; the New South
<http://www.virginia.edu/history/faculty/wthomas.html>

Allen Tullos, Emory's American South project
http://www.southernspaces.org/edi_boa.html

John Willinsky, University of British Columbia
<http://www.lled.educ.ubc.ca/faculty/willinsky.htm>
<http://www.pkp.ubc.ca/>

Also in attendance at the June 2005 meeting from the DLF project:

Christie Hartmann, DLF
David Seaman, DLF
Michael Furlough, University of Virginia
Sarah Shreeves, UIUC
Tom Habing, UIUC
Kat Hagedorn, University of Michigan
Martin Halbert, Emory University

Purpose of the Meeting

The meeting achieved several interlocking goals: introduced the members of the Panel to each other and to the librarians taking part in the grant activities; allowed the librarians to explain what harvestable metadata is, how OAI works, and what its potentials for digital scholarship are; allowed the scholars to ask questions about their role in the project; and allowed the librarians to test assumptions against a sample audience and to receive feedback fairly early in the grant's timeline.

The bulk of the first half of the meeting was taken up with a series of discussions and demonstrations of existing OAI tools and prototypes, and to an explanation of where OAI has come from and where we hope to take it as a library service. The University of Michigan's OAister, Emory's Metadata Migrator, UIUC's Experimental OAI Registry, the CIC OAI collection, and The University of Waikato's Greenstone were all demonstrated.

The second half of the meeting on the following day started with a discussion of the OAI Best Practices work, MODS, and DLF Aquifer. There was early on an observation from a scholar that many of their Web projects are going into their second generations and this may be a good time to influence scholars to add metadata and use services such as OAI harvesting. It was also noted that there is an inherent tension between the individualism that drives much humanities work and the consistency across projects that a service based on harvested metadata requires.

The meeting gave the attendees an opportunity to ask a series of basic, clarifying questions:

Q: How is an OAI service different to a library catalog (OPAC)?

A: A wider range of electronic material will be included – and a much narrower range of print material. The OPAC may have a collection-level MARC record for a body of archival material, for example; an OAI services may have simple item-level records.

Q: What is an OAI record a record of? A print item? An electronic one? There was some confusion over why an OAI record exists when there is no digital surrogate attached to it.

A: Typically of electronic items, but not at all limited to that technically.

Q: What is the purpose of OAI?

A: To enhance access to materials otherwise hard to find, in order to foster better scholarship and teaching.

Q: How is an OAI service different from Google?

A: This led to a long discussion in which we referred to the ability of OAI to use controlled vocabularies and thesauri; tailor the delivery to our users' needs; and expose content held in databases that Google misses (the so-called Dark Web).

Q: Does the appearance of a digital library item in an OAI record and subsequently in an OAI service give us any assurance that the link will be permanent?

A: It was no surprise to see this manifestation of the scholars' concerns with the lack of permanent URLs for the online library resources on which they rely. There was predictable disappointment that the metadata contained a URL that was not necessarily permanent, and there could be no explicit assurance given by OAI service providers that the content will still be there. This led to a discussion of caching the content linked to from OAI services, as Google does for content in its search engine.

Findings

The discussions over the course of this meeting elicited much feedback on the possibilities of OAI from the scholars' perspectives, and on the shortcomings of the current implementations. These notes gather the observations and comments thematically, but not strictly chronologically.

Function

Much of the conversation centered on discussions of how the scholar works, and would like to work, and how OAI may be able to aid and support those working habits.

There was very high value placed on being able to find quickly archival material on a subject. This was seen to be the single best potential for OAI services of the sort that we demonstrated.

Gathering: Lots of interest in the creation of virtual collections, a personal book bag (gathering) feature, and the ability to make collections that can be shared with a class, or with a scholarly community. A common desire was to be able to capture metadata records and manipulate them locally, perhaps in a citations program such as EndNote. This is a version of the desire to be able to capture the content itself into a local “filing cabinet”, personal digital library, or annotations tool. This desire to gather, hoard, and annotate pre-dates the Web (it recalls John Unsworth's discussions of basic behaviors and academic practice of what he calls “scholarly primitives”), but is all the more understandable in light of the concern that a resource may not be there when next you return to its URL (“link rot”).

Metadata alongside the data: There was a clear desire to have the metadata available at the point that one views the data object. At present, the OAI record typically includes a URL to the object, but once you follow the link the metadata disappears—the user cannot access it from the library page on which the resource exists. We see the metadata as a finding tool; the scholars clearly saw its value as a reference or citation once the object was on the screen.

Alerting and Profiles of Interest: High value was attached to any sort of alerting services if one can be specific enough about what one wants. Problems seen: too many alerts if too general; no alerts if too specific.

Multi-searching: If one gets no result in an OAI search, the OAI service could send your search on to Google. This led us to look at Amazon's A9 and to discuss briefly the pros and cons of such rich but crowded displays.

Browsing: There was a good deal of interest in being able to browse by subject and by collection, as well as being able to find items via search.

Context

There was a need for the services to be much clearer about whether a record was for an item or a collection, and to have better connections between the two. Can you easily see the collection in which an item belongs?

There was considerable confusion in how the search results are sorted when they come back—the main desire seemed to be to sort item results first by the collection they belong to, and by their subject, and only then by institution.

One of the acknowledged strengths of OAI over the Web search tools is that you have fields by which to limit a search—for example, subject, date, or author. However, when not every record in an aggregated collection has -- say -- subject terms, it was thought to be helpful to know how much or how little of the catalog you are searching if you invoke a limiter. That is, within OAIster's 5 million records, how many are you searching if you use a "subject" limitation?

Authority

How often is a record for an item created by someone other than the author/owner/creator of the resource? One scholar found a record for one of his own Web resources in both HUMBUL in the UK and in the DLF collections registry, neither of which he had had a hand in creating or knew anything about. Can you always tell if it is a third party who is the source for a collection or item description? Are these as trustworthy?

Appearance and Usability

There was a strong feeling that the services we demonstrated were too "library-focused"—they gave too much priority to the name of the institution from which an object came—this was seen as good PR for the library but not a particularly useful thing for the scholar to know (at least, not of prime value earlier in the searching process). The item and its collection are of prime importance, not the institution that holds the item.

The interfaces did not have the look of something that is easy, approachable, and useful—they looked dense and ponderous (like finding aid services, said one scholar who had obviously had a bad experience on a special collections Web site).

As seen elsewhere, we also saw a clear desire for the search tool to mimic Google in its initial search screens and in the layout of the results returned—simple search first, and then advanced search one layer down; simple display first, and then full-record display, with record fields names visible. A suggestion was to put in the users' control the decision of whether results came back in a short form or in a full record view initially.

The use of a thumbnail image in the metadata record and the "text grabber" function Tom Habing described were considered to be very useful indeed, which are part of the CIC OAI project. The search interface with the thumbnails is the "Keyword search" located

at: <http://nergal.grainger.uiuc.edu/cgi/b/bib/oaister> . Only 10-20% of the records have thumbnails at present; to see them in action try any of the following searches:

"steel mill"
"war damage"
"illinois river"

CIC also uses thumbshots, which are shrunken screen captures of web pages, in their "Collection Browse" located at: <http://cicharvest.grainger.uiuc.edu/colls/collections.asp>

Several times we heard comments about the importance of the name of the service, marketing it to scholars, and explaining how this is different to a mini-Google. We were encouraged to stress the fact that this material is largely scholarly and often comes from archives, special collections, and museums. That is, it collects together material that these scholars are interested in and that they often have a hard job finding.

Tools

Finally, there was considerable interest in tools for harvesting OAI and for creating it for their own material.

Most Valued Services

As an exercise late in the meeting, we summarized the features that they felt collectively were of most utility—a wish-list, untempered by our ability to deliver them all. The first two, when put to a vote, were considered most valuable of all.

- a) The ability for the user to download the metadata in a fielded or tagged format, for use locally (in a bibliography, personal library tool, in EndNote, etc)
- b) Tools to aid institutions—especially those with rich holdings but limited technical expertise—to create sharable metadata
- c) Thumbgrabber / textgrabber services to enrich metadata records with snippets of the content they describe
- d) Metadata conversion/migration tools for individual users
- e) Multi-search of OAI repositories alongside Google, etc.
- f) The ability to find an item and invoke a “More Items Like This” service
- g) Alerting services
- h) The ability to search abstracts where they occur

- i) Annotation by user (Wiki-like?)
- j) Clustering and categorization by subject for browsing/searching (“search within”)
- k) Normalization
- l) Display collection name/institution name for item

Next Steps

The meeting concluded with a sense that we had learned a lot about the scholars’ reactions to services and digital library protocols that we were familiar with but which were largely new to our scholarly advisors. It also underscored how useful it is to have such a panel of experts as a part of the planning and development of a research and demonstration grant such as this one, and not simply as commentators on a finished service built “on spec.”

The Scholars’ Advisory Panel remains in place (along with a Technical Advisory Panel) for the duration of the grant and we will be convening them again by phone and in person as we move into the second half of the project.

In the meantime, we are analyzing their comments and service “wish lists” to see how quickly we can offer up prototypes of services for them to react to that are guided by their sense of what would be most useful. There are over 30,000 OAI records for four DLF member organizations that already have MODS metadata in them, and it will be with these richer records that we will test out the scholar-driven OAI prototype services.

Selected Resources Referred to in the Course of the Meeting

The Open Archives Initiative (OAI)

www.openarchives.org

The Distributed Library: OAI for Digital Library Aggregation. IMLS 2004 National Leadership Grant for Libraries, Research and Demonstration. The current DLF grant.

<http://www.diglib.org/architectures/oai/index.htm>

OAI-based Services

Grainger Engineering Library at University of Illinois at Urbana-Champaign: Search OAI Information in Engineering, Computer Science, and Physics <http://g118.grainger.uiuc.edu/engroai/>

OAIster: an example of an OAI service that aims to include every OAI record that points to a digital object <http://www.oaister.org/o/oaister/>

Index to the DLF OAI Portal: under development for this project. A portal for all OAI records (items and collections) from DLF institutions

<http://www.hti.umich.edu/cgi/b/bib/bib-idx?c=imls>

<http://www.hti.umich.edu/cgi/b/bib/bib-idx?c=imls;page=simple>

An Experimental OAI Registry at UIUC: of use principally to builders of OAI services <http://gita.grainger.uiuc.edu/>

Examples of OAI records from the Library of Congress: see the raw material that OAI service providers gather

http://memory.loc.gov/cgi-bin/oai2_0?verb=ListSets

The Southern Digital Archives Conspectus: This website documents the library- and museum-produced, open access digital collections currently available on the topics of history, literature, and culture in the U.S. South from the Colonial Period (beginning 1605) to the present. <http://southconspectus.library.emory.edu/>

The CIC OAI project: Access to digital library resources of major Midwestern universities (CIC -- Committee on Institutional Cooperation).

<http://cicharvest.grainger.uiuc.edu/>

Other Services

Community-built resources

Wikipedia: an example of a community-built, Wiki-based resource

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

Multi-Search

A9: Amazon's multi-search tool that allows one simultaneously to search Amazon, Google, and other services. <http://a9.com/>

Mass Digitization

Google Print (incorporating Google Library)
<http://print.google.com/>
<http://print.google.com/googleprint/about.html>

Directories of resources

Humbul Humanities Hub
http://www.jisc.ac.uk/index.cfm?name=collections_humbul&src=alpha

Tools

TAPOR: digital tools for the humanities
<http://tapor2.mcmaster.ca/TaporMain/portal/portal>

Greenstone: digital library in a box, with OAI support <http://www.greenstone.org>

Other Resources

Martha Brogan, *A Kaleidoscope of Digital American Literature*: An exploration of what digital resources offer to the study of American literature.
<http://www.diglib.org/pubs/brogan0505/>

Library of Congress Authority records: An authority record is a tool used by librarians to establish forms of names (for persons, places, meetings, and organizations), titles, and subjects used on bibliographic records. Authority records enable librarians to provide uniform access to materials in library catalogs and to provide clear identification of authors and subject headings. For example, works about "movies," "motion pictures," "cinema," and "films" are all entered under the established subject heading "Motion pictures."
<http://authorities.loc.gov/>

Digital Promise Coalition's Digital Opportunity Investment Trust
<http://www.digitalpromise.org/>