

Members Council Digital Libraries Interest Group

Minutes

Monday, October 21, 2002

OCLC Smith Conference Center
Meeting Room 2
9:00-11:00 a.m.

Agenda:

9:00-9:30 Introductions.

9:30-10:00 Discussion—What is a Digital Library?

Prereading:

Lynch, Clifford. 2002. "Digital Collections, Digital Libraries and the Digitization of Cultural Heritage Information." *FirstMonday* 7,5(May 2002). Available online at http://www.firstmonday.dk/issues/issue7_5/lynch

10:00-10:15 Scope & role of DLIG: <<http://www.oclc.org/oclc/uc/feb02/rig.htm>>

10:15-10:30 What topics should the DLIG group discuss this year?

10:30-11:00 Presentation: OCLC Research (OR) Digital Library activities (L. Dempsey)

Members attending:

Kenneth Bierman, Diane Dates Casey, Ann Marie Clark, Lorcan Dempsey, Paul Gherman, Shirley Hyatt, Tammy Rabideau, Alvaro Quijano Solis, Jay Starratt, Sandra Yee, Michael York

Chair Paul Gherman convened the meeting just after 9 a.m. with a round of introductions. Members briefly described the status of digital-library (DL) activities at their institutions. Gherman remarked on the amount of digitization going on at member libraries, linking that reported activity with the *FirstMonday* article by Clifford Lynch that had been distributed to members before the meeting. This initiated a general discussion on digital libraries, which was the agenda item for the first part of this session.

Other than recognizing the current importance of DL activities, there were no specific decisions or resolutions made during this portion of the session. Major points discussed include:

- the variety and importance of current DL initiatives

- the idea discussed by Lynch of connections between collections generally was considered insightful, if perhaps just over the horizon of some practitioners.
- a recognition of the importance for OCLC to be involved in DL activities
 - important for OCLC & library community to clarifying the various roles surrounding DL activities, and for OCLC to choose the appropriate one(s)
- Open Archives Initiative, its Protocol for Metadata Harvesting, and related issues:
 - who currently is building harvesters?
 - National Science Digital Library (NSF funded)
 - Urbana/Champaign for IMLS
 - OAIster at Michigan
 - various activities in the UK
 - sustainability of harvesting activities
 - Will people pay to have their metadata harvested, or for a resource made up of harvested material?
 - Probably not now, but maybe in the future.
- Harvesting, serving up databases for harvesting, and providing a variety of modular services to others involved in these activities.
- “Moving collections up” and the idea of interpretation
 - perhaps not a role for OCLC, but rather for member institutions
 - on the other hand, there may be a role for OCLC to help libraries produce such material, organize & store it, & make it available. As we enter this space, libraries do come closer to archives and museums. We’ll have to work with them to work out roles. OCLC, perhaps, should make it possible to package, but not do the packaging. (although some may disagree...)
- as Lynch points out, we don’t really know the user community & how resources get used
 - although we are becoming aware that many materials previously considered to be less important are becoming heavily used when digitized (e.g., theses & dissertations).

Gherman then moved on to Lorcan Dempsey’s presentation re. digital libraries, available online at < <http://www.oclc.org/oclc/uc/oct02/interestgroups/dempsey.ppt> >.

- Dempsey stated his goal as wanting to contextualize thinking re. digital libraries within OCLC, and presented a grid currently under development in OR, for use in considering DL developments and opportunities
 - Across the top is “Stewardship” (the exact word is still under development)
 - materials on the right under this heading are subject to low level of curatorial attention within a lib or archive.
 - items on the left are the opposite
 - the other dimension (vertical) is one of uniqueness... things above the line are less unique (there are a lot of them)... things below are rare or “unique-er”
 - the grid can be used to characterize a range of resources and uses.

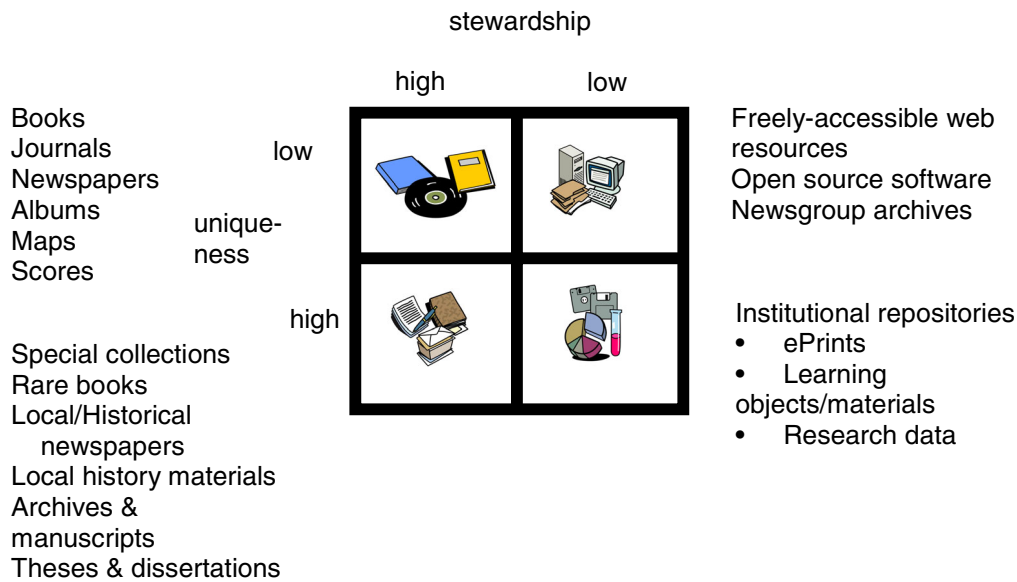


Figure 1: Grid from Lorcan Dempsey's presentation on Digital Libraries.

Examples of items in each quadrant:

I (upper left): heavily collected, but not unique. increasingly seeing electronic materials that are licensed. largest part of library investment & OCLC services are here... but where is the largest part of users' interest?

II (upper right): freely accessible web resources may be cheating somewhat... in a cataloging model, these do exist in many places by virtue of the cataloging records.

III (lower left): the domain of special collections & rare books, archives, etc. Theses & Dissertations also are here; much of the earlier discussion of digitization occurs here.

IV (lower right): materials typically subject not to great formal attention, but unique, including institutional data, research data and course materials; institutional repositories: a curated environment within institutions for such materials.

Digital library issues typically emerge below the line. The services OCLC provides largely cater to the top left-hand quadrant. What types of services are relevant below the line? Clearly, digitization & preservation support. Most of our other activity does focus in top left-hand quadrant.

We have quite a bit of apparatus for streamlining access to the commonly available resources. The question is, what services should be available below the line? (Harvesting & making freely available fits in here.)

Discussion of whether people feel this is a plausible way of capturing an aspect of the resource space we're operating in, & to relate some of the previous discussions to what's going on at OCLC. ePrints, research data, all will be the special collections of the future, but if you look at the current conversations regarding scholarly communication (see slide), it's about moving it down (below the horizontal line)... however, there's also talk re. moving some material back up... you want to license some things & move them back up over the line.

The question for OCLC is, should we be aggregating access to that type of data through OAI and other access routes?

After Dempsey's presentation, committee chair Gherman turned the conversation to the scope & role of the Digital Libraries Interest Group.

Dempsey and Shirley Hyatt suggested that since the charter was developed by a different group (albeit with some overlap), it would be good to reexamine & affirm or modify it.

Dempsey indicated that the authors intended for the group to take a broad view of digital libraries, addressing such questions as how to create environments in which OCLC can provide useful services for providing access to content from all quadrants of the grid—as well as preservation services—and to get input about our activities & where we might refocus our activities.

Gherman mentioned that the Lynch article distinguishes between a collection and a library, and asked whether we should distinguish ourselves from the digital preservation group, since we're not necessarily about digitizing & preservation, per se.

Jay Starratt recommended focusing on OR activities. Dempsey indicated that the division is currently looking for a researcher to work closely with Digital & Preservation Resources, but other issues are under consideration, as well.

Gherman agreed that the committee's role is to provide support for OR.

Dempsey: "If you take a broad view of digital libraries, most of what we do would fall into that area."

Hyatt: One of our concerns last year was how to make this group more interactive, & how it might better meet your needs as well as OCLC's.

Gherman: One of the problems is that much of what Lorcan talks about is pretty far out there for most of us... removed from the day-to-day [but] we need OCLC out there anticipating what's coming, focusing our attention on it, and developing services for it.

Hyatt asked whether it would help to focus each meeting on one aspect of digital libraries, and whether it was useful for the group to have "homework" between meetings.

There was general assent that this is a good approach, although one member expressed the opinion that it would be good to balance between daily practicalities and longer-range considerations, so as not to lose sight of the overview.

Ann Marie Clark commented: “Even though we’re mired in the day-to-day, it’s good to know what you [OR] are reading.”

Tammy Rabideau termed it “a validating luxury.”

Dempsey indicated that he wants to consider more what we might do to support the interface with learning-management systems, mentioning Pat Pitkin & her work with these issues.

Alvaro Quijano Solis encouraged a more international flavor; Gherman and others urged consideration of image, audio, & video content... “putting it all together,” including dynamic, rather than just fixed, content.

Hyatt reminded committee members that they can share new ideas between meetings by posting to the committee’s e-mail list.

The session ended with several comments reaffirming the need for the committee to stay connected with what OR is doing, and the suggestion that OR make available a list of current activities.

Tuesday, October 22, 2002

Marriott Northwest Hotel
Columbus Room
8:30-9:45 a.m.

Agenda:

- 8:30-9:15 ARL Portal Vision Statement; How should OCLC fit into this?
<<http://www.arl.org/access/scholarsportal/may01rept.html>>
(Paul Gherman, Lorcan Dempsey)
- 9:15-9:35 Presentation/Discussion: Management Decision Support Statistics
(Lynn Connaway)
- 9:35-9:45 Wrap up and housekeeping

Members attending:

Ken Bierman, Diane Casey, Ann Clark, Lorcan Dempsey, Paul Gherman, Shirley Hyatt, Alvaro Quijano Solis, Tammy Rabideau, Jay Starratt, Sandra Yee, Michael York.

Committee Chair Paul Gherman convened the meeting at 8:32, and turned directly to Lorcan Dempsey for his presentation on Web portals and the ARL Portal Vision Statement.

< <http://www.oclc.org/oclc/uc/oct02/interestgroups/dempsey.ppt> >

Lorcan Dempsey began by posing the question, “What is meant by ‘portal’?” His answer: “how the library mediates between user & resources in a network environment.” Resources are increasingly diverse; you need a variety of services “in the middle” in order to hide the diverse sources from users and present integrated interface (GUI, query, etc.), as well as to provide access to appropriate copies, etc. (SFX; & profiling).

For the system to work well, we increasingly need metadata about a variety of things: objects, rights, policy, collections, services, users, institutions, & schemas.

Portal application includes cross-searching, ILL; the actual applications that interact across the range of things.

Others will provide the portal application; OCLC is positioned to provide at least some of the utility services that would be useful across portals, regardless of who produces the application.

Portal Utility Services are being developed at OCLC in Cooperative Discovery Services, under Frank Hermes. [The ILL policy directory is more developed than other OCLC utility services.] OR, in particular, has the Metadata Switch project, which includes name authority, knowledge organization services, and schema translation services. We’re also working with Cooperative Discovery Services.

Paul Gherman: The origin of directory services would be at the institution?

Dempsey: I think there would be a mix, with ILL policy data coming from the institution and collection information coming from a mix. This is all very preliminary, but we’re thinking there would have to be some shared data entry and maintenance to make it all work. Within the metadata area we’re considering a rights and resolution service; this would be a central place where people could manage a variety of their rights-related data. Re. identity service—there has been some suggestion that OCLC could provide authorization services, but we’re not pursuing that as a 3d-party service.

We’re doing directory, rights & resolution, exploring potential for terminology; but at the moment we’re not in the identity management space.

Slide: OCLC: planning / research

Gherman: How can Members Council influence or help with this development? I’m in an ARL; I have heard of the portal project, but haven’t focused on it. What’s OCLC’s role, and how do we help them play that role?

Dempsey: If people developed a shared architectural view, this would help advance the discussion with providers. There exist a variety of mental models, which confuses the general discussion. It would be good to have a shared architectural perspective, what you'd like to get from a vendor, and what you'd like to provide; then you could have an interesting discussion regarding roles.

Gherman: I'm never sure what people are talking about when they say portals. ARL initially seemed to be talking about a portal that is to higher education as Yahoo is to the world, but they seem to have expanded considerably beyond that into services and other things.

Dempsey: Although I do not have any intimate knowledge of ARL, they do seem to be talking about a common interface for A&I and other resources, and perhaps, to surface resources elsewhere, e.g., Learning Management Systems (LMSs).

Sandy Yee: There also are portal-development activities on campus, seeking to integrate all of an institution's resources, including beyond the library, e.g., the LMS

Dempsey: Right. We're talking about integration, but the real integration is with what other systems exist & how people use them. Assent from Sandy, who commented on the high degree of interest and activity going on around LMSs, enterprise portal, student portal, administrative systems, and library portals.

Gherman: I'm concerned that individual institutions don't have much of a voice in the conversation; we don't know what it is or where it's going on. When I think about yesterday's digital libraries conversation & this one, I'm sure they relate, but I don't know how.

Dempsey: Well, the digital library would be one of the blobs behind the portal interface. But, you're right; we don't have a venue where we can develop a shared view

Gherman: I'd think the content providers also would want to provide their products in a form that can integrate with this.

[Some assent, but few people are talking.]

Bruce Newell commented about standardized or modularized protocols to allow for common tools for working with all of these systems, each of which currently is very different.

Dempsey responded about OCLC's participation with standards-development activities.

Yee: Has OCLC defined a way to search FirstSearch & WorldCat (WC) together?

Dempsey: You can do that... that's the first step... a distributed query... then you run into an authentication issue, then service description, etc. You have to have a lot of data available about the resources you want to interact with.

Yee: But we already have all that information available.

Dempsey: Right; much of it.

Gherman: Let's move on... but this is important enough that we should come back.

Lynn Connaway presentation: Management Decision Support Statistics

< <http://www.oclc.org/oclc/uc/oct02/interestgroups/connaway.ppt> >

A variety of names have been proposed for these projects; right now they are being called "Management Intelligence."

That heading includes:

- Comparative Collection Assessment
--a 2-phase project to provide data for member libraries for decision-making.
- an ILL Study, using OhioLINK data
--Data will be collected this month; researchers hope to have the analysis by spring.
- ARL eMetrics and COUNTER
--COUNTER is an international group that seeks to work with vendors to standardize statistics. Vendors could be certified by COUNTER. Includes vendors, funding agencies, & libraries.
- User Behavior Study
--identify how print and electronic materials are used

Connaway invited comments and suggestions, including input regarding other aspects of this work that would be of interest.

Newell expressed interest in age differences, e.g., horizontal vs. vertical reading (screen v. paper flat on table)... for public libraries in particular, since they serve such a wide range of ages.

Dempsey: An important part of this for us is that we want to see how much we can extract from WC that would be helpful, initially for collection development at netLibrary, but then beyond that how useful it can be. WC; QuestionPoint; ILL, which must be viewed in context. We have a very rich resource in WC & we want to see what types of

things we might be able to extract from that to support mgmt decisions of a variety of types, both within OCLC & outside.

Gherman: It seems to me to obviously be of very great value. ILL is expensive, so information we have about less costly alternatives is potentially a very exciting prospect.

Yee: We're participating in the PDA project from netLibrary, so we're letting our patrons help us develop our collection (don't buy until it's been clicked on twice).

Connaway commented that she had recently had an opportunity to observe user behavior in a controlled setting. The experience brought home the importance of looking at individuals of different backgrounds, ages, academic disciplines, etc. She saw undergraduates being "quick & dirty," "getting in then getting out," while faculty read pages and pages of material before deciding whether or how to use an item.

Gherman: A wonderful service from OCLC would be a notification service to say you have [one of] the only copies of this item: take care of it!

Q. from unidentified woman: Is there some way to get a global picture of what you've acquired for a year & how does that relate to your peers... is your vendor really doing a good job for you? Could provide staff savings.

Dempsey: Right... what collections are like mine; what should I be comparing to? How useful is WC for this type of question? It's less so for acquisitions, due to cataloging lag. Someone internally has been working on this, and the suggestion is that one can expect to have all the books cataloged from a given year, within 8 years (since libraries sometimes hold on to original cataloging items in hopes someone else will do it first.)

Gherman: We're almost out of time:

1. Regarding "Research" in the Interest-Group title... mind if we put it back in?
Response: OK to do.
2. Reminder to send comments to Lorcan Dempsey, Paul Gherman, Jay Starratt, or post to list.