

Catalogers Group meeting
August 8, 2002.

Attending: John Riemer, Louise Ratliff, Janice Matthiesen, Beth Feinberg, Joan LoPear, Valerie Bross, Angela Riggio, Nancy Norris, Sharon Benamou, Luiz Mendes, Jeff Morehead (recorder)

I. First UC Digital Library Forum

John discussed the 1st UC Digital Library Forum held on Monday, the 5th of August, 2002, at a hotel near LAX. He described the meeting as divided roughly into two halves. The morning section was about "ERDb"-like portal systems, and the afternoon session was a series of presentations on digital projects around the UC system.

Dan Greenstein, currently executive director of the CDL and former director of Digital Library Federation, gave the keynote, in which he anticipated a trend of digital library work emerging "fromof the ghetto." The "ghetto" refers to digital library work that is typically being done in separate units in libraries. The metadata created in those projects would be integrated with the contents of the OPAC, or at least searchable simultaneously with the contents of OPACs. Greenstein also said, "We need achievable standards and implementation guidelines--not pure, not perfect, but working." Mr. Greenstein also discussed the high costs of collaboration, but suggested that at the national level tools and services might be developed for application at the harder-to-afford local level.

The term "portal" was noted as having many meanings, ranging from "web starting place" to "super discovery tool."

Motivations for the creation of UCSD's Sage include overcoming the inefficiency of having to maintain descriptions for the same e-resource on multiple static web pages and the desire for an academic alternative to commercial search engines. Sage enjoys a high level of buy-in. Sage supports ERDb-like browsing by subjects and types. A relational database lies at the bottom of the search engine. E-journals reside in the Roger OPAC, but not the free monographic resources like websites. Title and URL are the only 2 required fields in Sage! Anyone can edit any record, but people are expected to consult first. UCSD has found that dynamic pages can be slow, so they regenerate static pages nightly.

Dora Loh and Andy Kohler presented "ERDb: Portal or-- Pig-in-a-Poke?" It is the only database aiming to house licensing metadata as well as to support public discovery. It was said the ERDb has an interesting set of deliverables, ranging from a data element dictionary to usage reports, policy issues, email capability etc. Our product was also noteworthy for the degree of emphasis on synchronizing data with the OPAC.

UCR's Infomine is a "librarian built" "virtual library." While SAGE and ERDb contents are selected by librarians for an academic audience, Infomine is not as discriminating. Containing 25,000 records, some of them date back to 1994, and some have come from non-UC sources such as Cal State Universities and Wake Forest University. The metadata of some of the records is Dublin Core based. Infomine is interested in the development of harvester tools, including automated LCSH and LCC assignments. John later learned that Infomine has two grant supported Ph. D. programmers. Infomine is currently obtaining records from Cornell and LC.

Dan Greenstein said that we need to develop objective cost/benefit criteria against which to evaluate portals like these.

The afternoon focus shifted to various five digital library projects taking place on UC campuses: The Alexandria Digital Library project at UCSB, CDL's "Counting California" and their "Online Archive of California," Berkeley-developed METS (metadata encoding and transmission standard), and a presentation by Curtis Fornadley concerning the Sheet Music Open Archives Initiative, much like the one he gave to Catalogers Group in April.

The scope of the Alexandria Project is anything spatially referenceable; it includes cartographic images and a gazetteer (2.8M bib records featuring 5.5M place names). It is navigable using Mr. Sid software.

“Counting California” has a scope of government produced stats in the social sciences; both metadata and the raw numerical data are searchable.

The “Online Archive of California” consists of finding aids to special collections, many of which lead to scanned images of the material.

METS is a means of keeping track of all parts of a digital object, tying the descriptive and administrative metadata to the digital content itself. This standard is being adopted and considered for use by a number of institutions including LC and the CDL.

At the end of the meeting there was round robin from all campuses, answering 3 basic questions: How have you organized in support of digital library activity (centrally or not?). What noteworthy projects are currently under way? What challenges do you face? Projects of note: UC elinks, Melvyl-T, visual resources infrastructure (CDL), campus infrastructure (UCB), Women Romantic Poets project (UCD), sheet music OAI (UCLA), Alexandria Digital Library (UCSB), a Chicano studies (also UCSB), e-journals licensing database and arts library slide collection of 25,000 items (UCSD), tobacco documents (UCSF). Other campuses are just getting started.

--Louise asked John to describe the attitudes of the various participants. John said he felt that everyone was very interested in and serious about the projects they are working on.

--Janice asked whether subsequent annual meetings would be invitational and whether there would be some way to access the discussions from more remote locations in order to create broader access. John said perhaps LAUC could be represented at future annual meetings.

--John added that there were discussions about the difficulty of getting anyone to agree on accepting one system. For example, could Infomine, ERDb, Sage, and CDL Directory efforts be merged in some way? John recalled a suggestion heard at the meeting that every campus should at least try to agree to a single subject vocabulary!

II. Highlights of August 7 CMC meeting

Next John discussed the CMC meeting yesterday. The ease with which a person could select dozens of e-resources for cataloging points up the need to (re)establish selection criteria for free resources. A new CMC task group, consisting of Ellen Broidy, John, and a couple of others will need to start by examining the 1996 Geyser [sp.?] report on e-resource selection. Cindy Shelton, Sara Layne, Norma Corral, and Barbara Schader were the on previous committee. The current abundance of “free e-resource material” has led to a reconsideration of selection criteria. Aspects of cataloging such what level of cataloging will be used, workflow, removal of Orion records when access to the resource disappears, will also be included in the new group’s charge.

John added that YRL Cataloging was complimented on how many web sites have been cataloged in the OPAC. He added that the Management Library has discovered to its regret that nothing of theirs has made it into the ERDb. Apparently this is a result of some of their material’s being licensed specifically for use in the Anderson School. Fewer than a dozen titles are restricted to that degree, it turns out. When ERDb is augmented with resources cataloged in Orion since last September, many Management titles will migrate.

Norma Corral sympathized with us for the lack of selectivity involved in acquisition of online federal documents (the Library is expected to provide access to everything issued).

--Valerie asked who was involved locally in determining selection criteria for “free e-resources.”

--Luiz noted that current selection criteria seemed to result in a criss-cross effect, especially in government documents areas, in which e-resources which were selected for cataloging were actually parts of larger sets of material, and that it was just not clear whether the whole set of sites should be cataloged, or the specific site, or whether the selector was aware that there was a larger, umbrella site.

--Valerie said that levels of granularity were a BIG concern for catalogers. She also warned that it was difficult to compare virtual resources with tangible resources, but that current selection policies seemed to do that.

--Louise noted that some web sites are spontaneously given addresses by the site engine which cannot be used to create a URL (or PURL); the site exists but access must be through an umbrella site.

--Luiz suggested that a management structure supporting the selection and cataloging of “free e-resources” be developed.

The final agenda topic, statewide cataloging policy for e-monographs, is deferred to a later meeting.

III. Announcements

John noted for the record that the Library of Congress has finally received our check for access to Class Web. An email containing the password and PIN information is expected by our Digital Acquisitions Coordinator imminently, and we should finally have access to it REAL SOON.

Jdm, 8.8.02