Heads of Special Collections -- HOSC

November 14, 2005
California Digital Library, Oakland

Present: C. Faulhaber (B), D. Morrison (D), J. Dooley (I, recorder), E. Lin (M), M. Conway (R), D. Tambo (SB, chair), C. Bunting (SC), L. Claassen (SD), L. Mix (SF).

Absent: V. Steele (LA).

CDL guests: Bill Landis (host), Steve Toub, Patricia Cruse, Rosalie Lack, Felicia Poe, Lynn Cameron.

1. **Interlibrary loan report (Morrison)**

   CDC endorsed the report and passed it up to SOPAG. We await word as to their response.

2. **Welcome (Landis)**

   Bill welcomed the group and apologized for the unavoidable absence of Robin Chandler, Director of Built Content. All presentations will relate to current CDL projects that may be of interest to HOSC.

3. **Online Archive of California (Landis)**

   With the departure of Adrian Turner, day-to-day management of OAC will be overseen by Rosalie Lack, assisted by Gabriela Montoya of UCSD (25% time). A backlog of problems is being addressed.

   OAC was transferred to CDL's standards-based common framework infrastructure last summer. The Calisphere interface is coming up in spring 2006, which will affect OAC. [N.B.: see also Calisphere agenda item below.]

4. **Interface Customization Tools (Toub)**

   Steve gave a PowerPoint presentation on the "skin-and-slice" techniques available for use by campuses for development of customized interfaces. CDL's view is that campuses should design front-ends for resources, since we best know our clients.

   The OAC home page does not link back to project-specific homepages at this point, but that capability could be useful for some users who find material in OAC of particular interest. Calisphere will give us a way to explore the relationship between virtual and tangible collections and their associated metadata. CDL will not assume maintenance of campus-based projects such as JARDA; the host campus must decide whether to add to each collection. How to keep such collections useful over the long term is a big question.

   All data in OAC is available for metadata harvesting by anyone via the OAI PMH protocol. CDL has licensed the Metalib product from Exlibris, which permits searching of multiple databases simultaneously.

Dooley gave a demonstration of UCI's soon-to-be-released SEAAdoc site, an NEH-funded project based in UCI's Southeast Asian Archive, and discussed how UCI partnered with CDL in its development.

5. **Web Archiving Project (Cruse)**

Patricia noted that she is available to visit campuses to speak on both of the projects that she addressed at the meeting.

CDL is eight months into the three-year National Digital Information Structure Preservation Project (NDIP), studying how to build a web archiving service that will enable a UC library to set parameters for the websites it seeks to collect. Sites focused on two topics are being gathered to build a testbed: political elections and Hurricane Katrina. Funding is principally from a $3 million Library of Congress grant. The Internet Archive is also a partner. CDL is costing out how to build such a service but hasn't yet determined costs of managing it on an ongoing basis. Metadata is a big issue: what level of data is necessary, and who should do the work?

Landis noted that websites are the "printed ephemera of the digital age" and therefore constitute key primary sources that must be collected. He also noted that archival collective description could prove a valuable model.

The duplication rate across archiving projects is unknown, but Cruse noted that true duplication is unlikely because there are so many different parameters, technologies, and approaches used in such projects.

6. **Digital Preservation Repository (Cruse)**

The DPR is meant to be a self-service utility built on tools that campuses will utilize to use the repository in a self-sufficient manner. The available functions are to submit, archive, and retrieve content. It is easy and flexible enough to be used without needing local programming support, but a Java toolkit is available for campuses wanting more sophisticated capabilities. Management reports are being developed. A campus can request dissemination of its archived content, but the DPR itself is not a user-access system or a passive backup system; it is intended only for storage of "curated objects." A glossary of terms has been developed to ensure understanding. Submission agreements are important to ensure that campuses understand their own and CDL's roles and responsibilities.

CDL's role is technical and administration oversight. Campuses can submit anything as long as they have the right to do so. One of three circumstances must pertain: material is in the public domain, the campus has copyright, or the campus has signed an agreement with the copyright holder permitting deposit. CDL knows that campuses will submit materials that don't meet these criteria; it is the campus's decision to do so. Campuses should therefore be seeking permission to archive a digital copy at the same time they request permission to digitize and make surrogate versions publicly available.

CDL is currently working with four campuses to move their first deposits into the production stage. Cruse, together with Dan Greenstein and Gary Lawrence, is also launching the Digital Stewardship Project to do an environmental scan of all digital data produced on UC campuses. The goal is to get a sense of the digital universe that is out there to be preserved.

7. **Round-robin discussion over lunch (All)**

   Mix raised the issue of digital management. Cruse suggested that HOSC try to articulate our needs for such systems so that CDL can determine whether it's possible to build something. Bunting asked why the Image Service can't subsume such needs. Landis noted that the Image Service is not METS-based because it is based on Luna's Insight system, but that CDL recognizes that in future this will be necessary. For example, CDL is looking at bringing its two large image banks, OAC and Image Service; the former accepts METS.

   HOSC members may want to review the draft revision of CDL's Guidelines for Digital Objects, which are going out soon.

   Landis described CDL's American West "dashboard" project, the goal of which was to use metadata to facilitate creation of hierarchical browsing. This is CDL's first attempt to harvest metadata and deliver it to users.

8. **Calisphere (Lack)**

   Calisphere will repurpose OAC content for use by K-12 teachers. Lack's position as Public Content Manager focuses on CDL's public service mission, as mandated by the Regents. State educational standards mandate use of primary source materials and drive the topics and concepts that will be featured. She will be reporting on the topics for which OAC lacks sufficient image content and may put out a call to HOSC seeking additional images. Teachers want only about ten images per theme to avoid being overwhelmed. They place high value on UC as a source of reliable information.

   Chandler sees the issue of how to populate Calisphere an important one for future discussion with HOSC.

9. **Cal Cultures (Lack)**

   The site has been mocked up and will be a subset of Calisphere, as will JARDA at some point. CDL wants to work with HOSC to revisit the concept of OAC as a destination site once Calisphere and Cal Cultures are available. Is there a way to incorporate finding aids into Calisphere? Or to point to them from Calisphere?

   Landis is evaluating topical clustering of OAC finding aid data, which could help with the problem of reliable subject searching of OAC.

Submitted by
Jackie Dooley, UCI