IMLS Digital Collections and Content
Grant LG-02-02-0281

Interim Performance Report 7
1 October 2005 – 31 March 2006

Submitted by Timothy W. Cole, Principal Investigator, and
Jenny Benevento, Project Coordinator
University of Illinois at Urbana Champaign
April 2006

Summary

The IMLS DCC Project has continued progress on goals during this performance period. With the extension of the grant for 2 more years (effective 1 October 2005) we have made changes in project staff and steering committee members. We have begun an update of our web services and interface, migrating from a research project-oriented site and look-and-feel to a more production-oriented public interface. We have begun work to integrate item-level and collection-level metadata searching and have created and are planning testing of a prototype of that new design. Our project has been actively disseminating results and reaching out to projects, providing advice on metadata design and implementation. Whitepapers covering most of the major results from the first 3 years of the project have been posted on our Website. Preliminary research on making collection and item-level metadata records useful in K-12 context is underway. We continue to interview participating projects and add new projects to the Collection Registry.

General Project Activities

Project Staffing

Two new research associates have joined the IMLS DCC Project. Oksana Zavalina replaced Ellen Knutson as co-PI Carole Palmer's Research Assistant for this project, and Kurt Groetsch is working with co-PI Mike Twidale on prototype interface testing. We also began working during this performance period with Geoff Ross, a masters-level from the Graduate School of Library and Information Science (GSLIS), who is doing a for-credit practicum with the Project, investing issues surrounding collection-level metadata record enrichment project for K-12 audiences. Finally, we gained new Steering Committee members Elizabeth Liddy, Paul Marty, Sam Quigley, and Katherine Wisser.

Website and Search Developments

During this period we redesigned our website and have focused more on publicizing it. Our previous homepage focused on the project and its research. In the redesign, we wanted to bring our end-user search and browse interfaces to the forefront. Several libraries now cite the IMLS DCC Website in their subject guides, and we have been mentioned in about a dozen library-related blogs.

We have plans to integrate the collection registry and item-level repository, while still preserving a collections only search interface. We completed a new test interface of an integrated collection registry and item-level record, incorporating a subset of metadata available in our production Website. This prototype interface is available at: http://cicharvest.grainger.uiuc.edu/heritage/collections.asp. To create this new interface we have consolidated the information from our collections and item-level databases into an integrated new database schema. Concurrently we are upgrading item-level metadata harvesting and processing workflow to include additional normalization and enrichment transformations.

Timeline

Our project's goals are on target to the new timeline included in the grant extension proposal. A test region of IMLS DCC records has been set up in GEM. Some IMLS DCC collection records have been ingested into the test region in order to affirm that GEM's XSLT transforms are working properly and the mapping of our schema to GEM RDF is successful. Ingestion of item-level records will soon begin.

We added the 2004 grantee collections to the Collection Registry and surveyed them about their projects for research purposes.

We have identified an initial set of LSTA grantees from Illinois and will add to records describing their collections to the Collection Registry in the course of the next 6 months.

Dissemination

We have continued talking with projects one-on-one regarding their project's adoption of metadata technologies. We also have worked with several students from the Graduate School of Library and Information Science in order to give them practical experience in the digital libraries field. Project staff have presented at several classes taught at the School as well. The IMLS DCC project staff and investigators have published and presented on the various research findings from the project in several forums, including:

IMLS DCC Project Overview (http://imlsdcc.grainger.uiuc.edu/MuseumInformatics.ppt) Presented by Jenny Benevento at LIS 490MUG Museum Informatics at the Graduate School of Library and Information Science at UIUC.

OAI-PMH in Practice: Lessons Learned from the IMLS Digital Collections & Content Project (http://imlsdcc.grainger.uiuc.edu/Cole-OAI-In-Practice.ppt) - Timothy W. Cole. Presented at Middle East Digital Library Workshop. Held at Bibliotheca Alexandrina, 15-17 January 2006.

[OAI-PMH repositories: quality issues regarding metadata and protocol compliance: Part II Shareable Metadata](http://imlsdcc.grainger.uiuc.edu/OAI4_Tutorial_1_URLs.htm#shareable) (http://imlsdcc.grainger.uiuc.edu/OAI4_Tutorial_1_URLs.htm#shareable) - Timothy W. Cole. Presented at CERN workshop on Innovations in Scholarly Communication(OAI4), Geneva, Switzerland, 20 October 2005.

Slides from our presentations are available on the About page of our website.

Other Forums

Sarah Shreeves and Tim Cole continue their participation in the OAI and Shareable Metadata Best Practices Working Group sponsored by the Digital Library Federation (DLF) and the National Science Digital Library. This working group, formed as part of an IMLS grant to the DLF, is developing best practices for DLF members and similar institutions. Our experiences from the IMLS DCC work heavily inform the contributions we are making to this group. Sarah is serving as chair and co-editor of this work. In addition, Sarah Shreeves has been appointed the chair of the Aquifer Metadata Working Group, part of an initiative of the Digital Library Federation.

Upcoming Publications and Presentations

Palmer, C., Knutson, E., Twidale, M., & Zavalina, O. (2006). Collection Definition in Federated Digital Resource Development. In *Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology*.

Steering committee activity

The steering committee met on February 16, 2006 at the Web-Wise 2006 convention. Four new members (Elizabeth Liddy (by phone), Paul Marty, Sam Quigley (by phone), and Katherine Wisser) joined the Steering Committee. We asked them for advice and input on our plans for the coming year. The agenda is available (password protected) at:
 http://imlsdcc.grainger.uiuc.edu/SteeringCommittee/steeringcommagenda2006.doc.

Research

Data collection

The second round of phone interviews that focused on collection identity in the registry and perceptions of harvested metadata was conducted in October-December 2005. Transcription and coding of all eleven phone interviews is complete and preliminary analysis is included in the paper focusing on collection identity that project team submitted in February 2006 for the American Society for Information Science and Technology annual meeting to be held in Austin, Texas in November 2006.

Six additional in-person interviews were conducted at the Web Wise Conference in Los Angeles in February 2006. These participants represented six projects from various types of institutions including academic libraries, academic center, media agency, and an art gallery. The focus of the interviews was on the challenges faced by digital collection creators in relation to collection definition, metadata creation and application, unexpected audiences, and changes in approaches

over time as their projects developed. Transcription of these interviews is complete and analysis of the data is ongoing. We are developing coding schemes and working through the intercoder reliability process.

The 2nd-round online survey of the 100 initial projects that responded to the first survey in 2003 was launched in February 2006. The focus of this 2nd round of survey is on tracing changes over time in the type of material in the digital collection, metadata schemes used, the intended audience, and other specifics about the digital collection and its technical implementation. The data collection is ongoing; some preliminary analysis has begun.

We are continuing to analyze the changes to collection level records in the Collection Registry made since February 2005. We are also analyzing the transaction logs of registry use to assess types of searches conducted. Our interest is in identifying the correlation among subject keywords used by registry searchers with the GEM subject scheme used in the collection level description and the other controlled vocabularies applied by the projects.

In the usability track, we have been focusing on evaluating and testing the next iteration of the item-level repository interface. Our approach is a series of formative evaluations continually informing the ongoing iterative design of the interface. This effort has consisted of two components: a detailed evaluation of several of the most sophisticated, widely known and popular digital libraries/federated search applications -- with an eye towards discerning best practices in the presentation of item-level search results and linking -- and usability testing of the most recent iteration of the search interface, currently under development. Work on the competitive analysis portion is nearing completion. We have obtained approval by the UIUC Institutional Review Board of our most recent amendment, so that testing with several representative user groups will begin in mid-May. We plan to focus on two constituencies of potential users: librarians and teachers.

Research Plans April 2006 through September 2006

Data collection for the 2nd-round online survey will continue until the end of April 2006. Analysis will proceed over the summer to further extend our understanding of changes in collection identity perceptions, audiences, metadata schemes and controlled vocabularies used etc.

We will also continue to analyze the registry record changes and intend to incorporate this analysis into our final revisions of the recently submitted ASIST paper. The analysis of the collection registry transaction logs will be completed and submitted initially as a GSLIS technical report. Web Wise interview analysis will be incorporated, along with other data, into a more developed paper on some themes of interest that emerged in the course of developing the ASIST paper.

Related Activities

We continue to hold the metadata roundtable study group, and have increased the frequency from bimonthly to weekly. We are very pleased that the number of the regular participants in the metadata roundtable has increased. Recent roundtable topics have included the issue of sets in collection description, a report on the Middle East Digital Library Workshop, and a examination of the Extension of the Analytical Model of Collections and their Catalogues. The website, which includes a full listing of the metadata roundtable topics and background readings, can be found at: http://www.isrl.uiuc.edu/~dcc/mdrt.html.