

IMLS Digital Collections and Content

Grant LG-06-07-0020



Interim Performance Report 2

1 April – 30 September 2008

Submitted by:

Carole Palmer, Principal Investigator
cpalmer@illinois.edu

Amy Jackson, Project Coordinator
(217) 244-7809
amyjacks@illinois.edu

October 2008

University of Illinois at Urbana Champaign

**Graduate School of Library and Information Science
Center for Informatics Research in Science and Scholarship
University of Illinois at Urbana-Champaign
501 E. Daniel Street, MC-493
Champaign, IL 61820-6211
(217) 333-1981**

**Grant LG-06-07-0020
Interim Performance Report 2
1 April – 30 September 2008
Submitted by
Carole Palmer, Principal Investigator
Amy Jackson, Project Coordinator
October 2008**

Summary

The past six months of the continuing IMLS Digital Collections and Content (IMLS DCC) project, “Next Generation Digital Federations,” were primarily engaged in research activities for both the Digital Collection Evaluation (DiCE) subgroup and the Collection and Item Metadata Relationships (CIMR) subgroup. Other activities included development of a history portal, contacting non-IMLS funded collections, and preliminary discussions for realignment of the IMLS DCC collection description application profile with the Dublin Core collection description application profile. These and other activities are discussed in detail below.

General Project Activities

Project Staffing

A new research programmer, Hong Zhang, was hired to assist with the metasearch and interface portions of the project. Hong began work on the project in August of 2008.

Digital Collection Evaluation (DiCE) subgroup

The Digital Collection Evaluation subgroup refined research results regarding use of the free-text description field in the collection-level metadata records and presented the results at the 2008 DCMI (Dublin Core Metadata Initiative) Conference in Berlin. Other research projects included a pilot analysis of matches to collection-level user search queries from the 2006-2007 transaction logs.

The subgroup also evaluated non-IMLS funded digital collections for complementary content and contacted several of these collections. The Digital Library of Georgia responded and provided IMLS DCC with 37 collection-level records which were integrated into the registry. The Autry National Center Collections Online and the New Mexico Digital History Project were also contacted and corresponding collection records were created and integrated into the registry. Other collections did not reply to initial contact attempts. Arrangements were also made to present at the COSLA (Chief Officers of State Library Agencies) meeting in October, 2008.

Collection and Item Metadata Relationships (CIMR) subgroup

The CIMR subgroup, led by Co-PI Allen Renear, held weekly meetings in the spring and twice a week in the summer. Graduate research assistants Richard Urban, Karen Wickett and Wu Zheng played key roles in this work. Other regular contributors included Tim Cole, Dave Dubin, Amy Jackson, Michael Twidale, Oksana Zavalina and other members of the DCC research areas.

During this period we refined our definitions of the first two categories (attribute/value propagation and value propagation) and introduced a third category (value constraint). Early results with special reference to applications for sustainability were presented at Digital Humanities 2008 (June 25-29, Oulu, Finland). A technical problem in the formalization of these categories was presented as a poster at the Joint Conference on Digital Libraries (June 16-20, Pittsburg, PA). The current version of the framework was presented at the International Conference on Dublin Core Metadata and Applications (September 22-26, Berlin, Germany). A paper describing the broader significance of our work was submitted to IEEE Intelligent Systems special issue on AI and Cultural Heritage. See the attached bibliography for citations.

In the current phase of CIMR work, we are applying the categories we have identified to elements in the Dublin Core Collection Description Application Profile. We will continue activities begun this summer testing CIMR categories against collection-level and item-level metadata found in the IMLS DCC repository. New framework categories and refinements to existing categories are being identified throughout these activities. Finally, we have begun to work with the Interface group in order to develop strategies for exploiting the framework to improve discovery and navigation between collections and items in the IMLS DCC repository.

Metasearch

Metasearch was integrated into the portal and will be available for on-campus usability studies and testing during the next performance periods. Targets include Academic Search Premier, America: History and Life, Scopus, and Google Book.

Interface development

The Interface Design subgroup spent the performance period reviewing literature, determining interface design challenges for a collection of collections, and assembling examples of web portals with similar design challenges. Identified challenges included communicating the scope of the portal as well as the differences between items and collections. These examples were contrasted with Google's approach of providing access at the item-level without contextual information. The literature review focused on how humanities scholars search for and use historical images. Weekly subgroup meetings will start during the next performance period.

Collection description application profile

The project team began discussing realignment of the IMLS DCC collection description application profile with the Dublin Core collection description application profile (DC CDAP). The IMLS DCC collection description application profile was created before the final draft of the Dublin Core Collection Description application profile was released, and significant changes have since been made to the DC CDAP. The IMLS DCC application profile will be aligned with the DC CAP by the end of the next performance period.

Registry/Repository maintenance

- 2007 NLG Grantees: IMLS identified twenty 2007 NLG grantees with digital content and these projects were contacted for inclusion in the IMLS DCC registry and repository. Twelve projects responded and eight completed IMLS DCC Survey One. Collection records were created for all eight projects, and six of these have been vetted added to the DCC interface. Follow-up with non-responding projects will occur during the next performance period.
- Georgia Digital Library: Thirty seven collection records were received from the Digital Library of Georgia and integrated into the collection registry. Fourteen of these collections were funded by LSTA grants.
- Complementary collections: The Digital Library Federation's American Social History Online project supplied IMLS DCC with 107 collection records, and these records were added to collection registry. Two other collections with complementary content were also added to the registry.

In order to accommodate non-IMLS funded resources, a new portal was developed for digital collections with significant content related to American history. At the end of the performance period the portal was in late development phases, and release of the portal should occur during the next performance period. This collection registry initially contained 279 collection records. The item-level repository will contain nearly 800,000 records after harvesting and integrating nearly 500,000 records from American Social History online and related items from the IMLS DCC repository.

The IMLS DCC collection registry, displaying only IMLS and LSTA-funded collections, now contains 224 collection records. All item-level metadata harvests were refreshed during the summer of 2008, and the item-level repository now contains 335,500 records.

Dissemination

Presentations and Publications

Amy Jackson and Myung-Ja Han. "Metadata Interoperability and CONTENTdm," presented at the Midwest CONTENTdm Users Group Meeting, Indianapolis, Indiana, April 29-30, 2008.

Allen H. Renear, Richard Urban, Karen Wickett, Carole L. Palmer, Dave Dubin. "Sustaining Collection Value: Managing Collection/Item Metadata Relationships," Panel on Sustainability, Digital Humanities Conference, Oulu, Finland, June 2008.

Allen H. Renear, Karen Wickett, Richard Urban, Dave Dubin. "The Return of the Trivial: Problems Formalizing Collection/Item Metadata Relationships," Poster, ACM/IEEE-CD Joint Conference on Digital Libraries, Pittsburgh, June 2008.

Allen H. Renear, Richard Urban, Karen Wickett, Carole L. Palmer, Dave Dubin. "Collection/Item Metadata Relationships," Presented at the DCMI Dublin Core Conference, Berlin Germany, September 2008.

Oksana Zavalina, Carole Palmer, Amy S. Jackson, Myung-Ja Han. "The Descriptive Substance in Collection-Level Metadata," Presented at the DCMI Dublin Core Conference, Berlin Germany, September 2008.

Upcoming/Submitted Presentations and Publications

Oksana L. Zavalina, Carole L. Palmer, Amy S. Jackson, Myung-Ja Han. "Evaluating Descriptive Richness in Collection-Level Metadata." In *Journal of Library Metadata*: accepted for publication.

Amy S. Jackson, Oksana Zavalina, Carole L. Palmer. "Cultural heritage institutions sharing metadata: How museums enrich our understanding of collections." Abstract submitted to Museums and the Web, Indianapolis, April 15-18, 2009.

Other project publicity

Anne Blecksmith, "Visual resources online: Digital images of primary materials on public Web sites," *C&RL News*, May 2008 (Vol. 69, No. 5). <http://ala.org/ala/acrl/acrlpubs/crlnews/backissues2008/may08/visualresources.cfm> (accessed October 8, 2008).

Sarah Perez, "Digital Image Resources on the Deep Web," *ReadWriteWeb*, May 14, 2008. http://www.readriteweb.com/archives/digital_image_resources_on_the_deep_web.php (accessed October 8, 2008).

Wikipedia contributors, "New Jersey Digital Highway," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=New_Jersey_Digital_Highway&oldid=243249809 (accessed October 8, 2008).

IMLS DCC created a public blog to highlight current project activities: <http://imlsdcc.wordpress.com>.

National Impact

IMLS DCC has had positive impact on a national level throughout the course of the project. The IMLS DCC collection description application profile was adopted by the DLF American Social History Online project, and characteristics identified for shareable metadata have been publicized by the Metadata for You and Me training seminars and the Digital Library Federation / Aquifer Implementation Guidelines for Shareable MODS Records (available at http://www.diglib.org/aquifer/dlffmodsimpementationguidelines_finalnov2006.pdf).

Future plans***October 2008-March 2009***

The site visit to the Illinois State Library, as described in our original grant proposal, will be replaced with a visit to the Chief Officers of State Library Agencies (COSLA) meeting in Jacksonville, MS, October 6-8, 2008. Anticipated outcomes from this visit include contact with state librarians and state-wide digital resource developers and inclusions of new projects in the IMLS DCC and history portals.

The project team will have a two hour conference call with Advisory Board members on February 20, 2009. The purpose of this call will be to update board members on current project activities and solicit feed about future directions.

Related activities

We continue to hold the metadata roundtable study group with a frequency of 2-4 times a month. Recent roundtable topics have included Open Archives Initiative Object Reuse and Exchange (OAI-ORE) standard: Use cases and applications; FRBR and music: Variations project; and Overview of CIDOC CRM conceptual reference model and how it relates to the Collection-Item Metadata Relations (CIMR) work. 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>.