IMLS Digital Collections and Content
Grant LG-06-07-0020

Interim Performance Report 1
1 October 2007 – 31 March 2008

Submitted by:

Carole Palmer, Principal Investigator
clpalmer@uiuc.edu

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

April 2008
University of Illinois at Urbana Champaign
Summary

The first six months of the continuing IMLS Digital Collections and Content project, “Next Generation Digital Federations,” were spent refining project planning, forming a project advisory board, evaluating current content, developing plans for a logic based framework for collection and item-level metadata relationships, and working through OMB clearance. Other activities included refining survey instruments, analyzing collection-level metadata elements, and identifying contacts for LSTA projects. The performance period closed with an advisory board meeting in Chicago on March 24 and 25, 2008.

General Project Activities

Digital Collection Evaluation
The first months of the grant period were spent evaluating the current collections to determine collection strengths and identify subject areas in which to expand the registry and repository with complementary content. Collections in American Social History Online (http://www.dlfaquifer.org/) were identified as complementary in nature, and will be added to the registry in the coming performance periods. Additionally, team members began initial investigations into sources of information to assist in identifying LSTA digitization projects.

Additional research analyzed collection records in the registry to determine substantive and purposeful information represented in the records. Findings show that the Dublin Core free-text Description field often provides more accurate and complete representation of subjects and object types than the specified fields. Properties such as importance,
uniqueness, comprehensiveness, provenance, and creator are represented, as well as other
vital contextual information about the intentions of a collector and the value of a
collection, as a whole, for scholarly users. The results demonstrate that the semantically
rich free-text Description field can provide a valuable source of data for enhancing and
customizing controlled vocabularies and that there is a need for collection-level metadata
schema that better represent holistic properties of collections.

**Collection and Item Metadata Relationships (CIMR)**

In the first phase of CIMR we are developing a logic-based framework of collection/item
metadata relationships that classifies metadata elements into varieties of convertibility,
with, when possible, associated rules for propagating information between collection and
item levels. This effort is led by Allen Renear and Richard Urban. Weekly coordination
and planning meetings began in November 2007 and are attended by 10-15 people from
the DCC project and the wider GSLIS community.

Focusing on the elements in the DCMI Collections Application Profile, we have identified
and defined our first two collection/item relationship categories: attribute/value-
propagation and value propagation. So far we have been working in first order logic and
focused as much on method as results, but in May we will begin using tools such as
Protege and languages such as OWL and moving more briskly through the elements of the
Profile. GSLIS is providing us with additional Research Assistant support for the summer:
Karen Wickett and Wu Zheng, both of whom are already part of the CIMR working group.

**Metasearch**

Candidates for the Research Programmer position in the library were interviewed during
this performance period. The position will be filled in the next six months and will
primarily assist with the metasearch and interface development aspects of the projects.

**Interface development**

Preliminary investigations into interface representation best practices and the use of
interactive visualization to identify data quality issues in federated collections have been
undertaken during this performance period.

**Registry and Repository maintenance**

OMB clearance for surveys administered during the previous grant cycle ended at the
completion of the previous funding. Documents for the current surveys have been
submitted to OMB for review, and should be completed around the end of May. Until
OMB clearance is received, no new collections can be added to the collections registry.
Harvests for the metadata repository were refreshed in October, and no new collections
have been harvested in this performance period.

**Advisory board**

In consultation with IMLS, a nine member advisory board was formed in January 2008.
Advisory board members are Martin Doerr (Centre for Cultural Informatics of the Institute
of Computer Science, Crete, Greece), Jeremy Frumkin (Oregon State University), Jonathan
Furner (UCLA), Michael Jenkins (The Metropolitan Museum of Art), Pete Johnston (Eduserve Foundation), Bill Landis (Yale University), Jenn Riley (Indiana University), Jim Scheppke (Oregon State Library), and Guenter Waibel (OCLC).

The first annual meeting of the advisory board was held on March 24 and 25, 2008, at the Illini Center in Chicago, IL. In addition to advisory board members, other attendees included Jean Godby (OCLC liaison), Rachel Frick (IMLS liaison), and UIUC project team members Tim Cole, Amy Jackson, Bill Mischo, Carole Palmer, Allen Renear, Sarah Shreeves, Michael Twidale, Richard Urban, and Oksana Zavalina.

Dissemination

Presentations and Publications
Carole Palmer and Oksana Zavalina presented “Discussion of the findings obtained in a content analysis on IMLS DCC Registry collection records” at the UIUC Graduate School of Library and Information Science Metadata Roundtable, January 23, 2008.


An article discussing the team’s analysis of changes in item-level metadata over time was accepted for publication in the *Journal of Library Metadata*: Amy S. Jackson, Myung-Ja Han, Kurt Groetsch, Megan Mustafoff, Timothy W. Cole. 2008. “Dublin Core Metadata Harvested through OAI-PMH” in *Journal of Library Metadata*. Vol. 8 no.1.

Upcoming/Submitted Presentations and Publications
Amy Jackson and Myung-Ja Han will present “Metadata Interoperability and CONTENTdm” at the Midwest CONTENTdm Users Group Meeting, Indianapolis, Indiana, April 29-30, 2008.


Other venues
The project received mention on several advisory board members’ blogs following the meeting in Chicago, including the blogs eFoundations (http://efoundations.typepad.com/efoundations/), Inquiring Librarian (http://inquiringlibrarian.blogspot.com/), and The Digital Librarian (http://www.digitallibrarian.org/Digital_Librarian/Blog/Blog.html).

Future Plans

April - September 2008
We expect final OMB clearance for the survey instruments during the next performance period. As soon as clearance is received, we will begin to contact 2007 NLG recipients and add their collections to the registry. Collections with metadata available through OAI-PMH will be harvested and added to the repository. Complementary collections from American Social History online will be identified and added to the collection registry as soon as surveys are cleared through OMB. Other collections whose developers have made contact with us about inclusion in the DCC will also be analyzed for complementary content and added to the registry as appropriate. Contact will be initiated with LSTA digitization projects and the registry will be populated with appropriate collections. Experimentation will begin with different views of the resource so that results can be limited to only IMLS NLG funded projects while providing the user with a clear understanding of search results. Additionally, the collection level metadata schema will be realigned with the Dublin Core Collection Description Application Profile.

Future research plans
A comparative content analysis of collection-level metadata in DCC and three other large-scale national aggregations of digital collections (American Social History Online, American Memory, and National Science Digital Library) will be conducted to start building an understanding of current practices in collection-level metadata application for subject access, with a special attention to free-text metadata. Additionally, the DCC web log will be analyzed to reveal distinguishing patterns in collection-level resource discovery. Interviews and observations of a sample of United States history scholars (a significant user group of the DCC project) will complement transaction log analysis and metadata analysis with an aim to assess user experience with and perception of the role of collection-level metadata in digital resource discovery.

Related Activities
We continued to hold the metadata roundtable study group with a frequency of 2-4 times a month. Recent roundtable topics have included Open Library project, BibApp project, RLG Programs Descriptive Metadata Practices Survey, the Library of Congress Working Group's Final Report on the Future of Bibliographic Control, and metadata quality visualization. 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.