

# IMLS Digital Collections and Content

Grant LG-02-02-0281



## **Interim Performance Report 8 1 April 2006 – 30 September 2006**

*Submitted by Timothy W. Cole, Principal Investigator, and  
Amy Jackson, Project Coordinator  
University of Illinois at Urbana Champaign  
October 2006*

**University of Illinois**  
**1301 W. Springfield**  
**Urbana, IL 61801**  
**Tel 217.244.7809**  
**Fax 217.244.7764**

**Grant LG-02-02-0281**  
**Interim Performance Report 8**  
**April 2006 – September 2006**

Submitted by  
Timothy W. Cole, Principal Investigator  
Amy Jackson, Project Coordinator  
October 2006

## **Summary**

During the past six months, the IMLS Digital Collections and Content (DCC) Project has continued to make progress toward stated goals and objectives. A second survey was sent to the initial 100 projects to track the evolution of digital projects, and we continue interviewing participating projects and adding new collections to the collection registry. A preliminary group of LSTA projects have also been added to the collection registry. A new search interface focusing on individual collections was tested during this period, and we continue to integrate item-level and collection-level metadata searching. Item-level metadata reprocessing and augmentation techniques were examined, with plans to implement these findings during the next performance period. As of September 2006, the IMLS metadata repository contained 245,012 records from 33 OAI-compliant NLG projects, and the collection registry contained records for 167 NLG digital collections. Our team also continues to publish and present findings from research performed as part of this project, and to provide advice on metadata design and implementation.

## **General Project Activities**

### ***Project Staffing***

A new project coordinator, Amy Jackson, was hired and began working on the project in September of 2006, replacing Jenny Benevento.

### ***Website and Search Developments***

A new interface featuring a subset of DCC collections was tested during the summer of 2006 (available at <<http://cicharvest.grainger.uiuc.edu/heritage/collections.asp>>). This interface shows both items and collections and allows grouping of item results by collection. We are currently discussing the usability of this approach and scalability to the larger collection registry.

We also investigated item-level metadata reprocessing and augmentation and plan to integrate these findings into the regular workflow in the next performance period.

### ***Timeline***

Our project's goals and targets continue to be met as scheduled on the timelines adjusted on previous interim reports. Work on metadata normalization, enrichment, and transformation continues on target, as does streamlining of processing and maintenance. Research into collection identity and metadata granularity continues through analysis of surveys and harvested metadata, and usability of new interfaces is currently being studied. Survey Two is being analyzed regarding development of expertise for collection managers, and methods for inclusion of LSTA data are being discussed. We have also continued work with GEM on ingesting DCC records into their test region (see Appendix Two for GEM interim report). We continue to provide information and assistance to NLG and LSTA projects regarding metadata creation and interoperability and OAI implementation.

### ***Financial Status Report***

The Annual Financial Status Report (Appendix One) has been forwarded to the IMLS Grants Administration office from UIUC's Grants and Contracts Office.

## **Dissemination**

The IMLS DCC project staff and investigators have published and presented on the various standards, protocols and research findings from the project in several forums.

In March 2006, Besiki Stvilia defended his doctoral dissertation "Measuring Information Quality" which incorporated analysis of the metadata quality in the IMLS DCC item repository. The dissertation will be disseminated through the university's new institutional repository. Other papers that led to the thesis work have been recorded in earlier reports. A recent paper that informed this work is:

*Shreeves, S., Knutson, E., Stvilia, B., Palmer, C., Twidale, M., Cole, T. (2005). Is Quality Metadata 'Shareable' Metadata? The Implications of Local Metadata Practices for Federated Collections. In H.A. Thompson (Ed.) *Proceedings of the Twelfth National Conference of the Association of College and Research Libraries*. (pp. 223-237). Minneapolis, MN. Chicago, IL: Association of College and Research Libraries.*

The project team's paper focusing on collection identity -- *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* -- was accepted in April 2006 by the review committee for the American Society for Information Science and Technology annual meeting to be held in Austin, Texas in November 2006. The paper was updated according to the reviewer's recommendations and the final version submitted in June 2006.

In June 2006, the analysis of the collection registry transaction logs was submitted as a GSLIS technical report -- *Zavalina, O. (2006). User Searches in IMLS DCC Collection Registry: Transaction Log Analysis. Technical Report UIUCLIS--2006/3+IMLS, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Champaign, IL [https://netfiles.uiuc.edu/zavalina/DCC\\_Project/Tech\\_report.doc](https://netfiles.uiuc.edu/zavalina/DCC_Project/Tech_report.doc)* -- and reported at the metadata roundtable.

Jenny Benevento presented *IMLS Digital Collections & Content and LSTA Grantees* to the Statewide Digitization Planners Meeting at ALA Annual 2006 in New Orleans, LA on June 24, 2006.

*The Chronicle of Higher Education* published an article describing the registry – Brock, Read (2006). “Federal Agency Unveils Database of Digital Collections from Museums and Libraries” in *The Chronicle of Higher Education* v. 52 (33), p.41.

## **Research**

### ***Data collection and analysis***

The second round online survey of the 100 initial projects that responded to the first survey in 2003 was conducted in February-April 2006. The focus of this 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 completed; the data analysis is ongoing to further extend our understanding of changes in collection identity perceptions, audiences, metadata schemes and controlled vocabularies used, etc. The data from this round of the survey, as well as from February 2006 interviews with digital collection developers conducted at the Web Wise conference and content analysis of the registry records, has been incorporated into the white paper on IMLS/NISO Framework of Guidance for Building Good Digital Collections.

In April-June 2006 additional content analysis of the IMLS DCC Collection Registry records regarding collection development policies, sub-collection delineation, and changes in Registry records made between January and June 2006 was conducted, with results incorporated into the final submitted version of the research team's collection identity paper that will be presented at 2006 annual meeting of ASIST.

We have continued to analyze the transaction logs of registry use to assess types of searches conducted and identify correlations among subject keywords used by registry

searchers with the GEM subject scheme used in the collection level description as well as with others widely used in the cultural heritage domain controlled vocabularies (e.g., Library of Congress Subject Headings and Art and Architecture Thesaurus). Analysis of about 500 user keyword searches in the collection registry made in February-September 2005 demonstrates a high level of subject searching made at the collection level (70% of the keyword searches represented subject-type search, including concept, class of persons, object, national/ethnic group, place, and event). We have also discovered the lack of semantic match between user queries and subject terms in controlled vocabularies. Only 2.6% of user search terms were matched in GEM subject scheme, while 22.63% were matched in Art and Architecture Thesaurus, and 71.3% in LCSH. GEM as the subject scheme representing only concepts seems to be incapable of meeting wide range of user queries in IMLS DCC collection registry.

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 completed and indicates that the latest iteration of the IMLS DCC site now is consistent with best practices in the field, addressing the challenges of not only providing multiple kinds of access to the resources available, but also helping new users to understand the nature of a federated collection consisting of both item-level and collection-level information. User testing with a focus on two important but distinct constituencies of potential users — librarians and teachers — was conducted between June-August 2006. A preliminary analysis of the data collected indicates that the latest iteration of the IMLS DCC interface supports and encourages collection-centric navigation and searching. A substantial portion of users, when presented with tasks, sought to identify the most relevant collections — based first on collection title and second on collection descriptions — in order to search within the those collections. This indicates the desirability of highly descriptive collection titles as well as concise collection descriptions that indicate collection coverage, as users frequently eliminated as irrelevant collections that contained relevant materials. Users generally found the presentation and content of brief search results sufficient, particularly when seeking visual materials. Despite the availability of full metadata records on the IMLS DCC site, users preferred to examine the item within its collection context. One area of ongoing concern is the usability of the "Relevant Collections" box on the search results page. Although the feature is potentially very valuable, its function is unclear to many users and, due to its placement and visual similarity to Google Ads, is frequently overlooked. Further data analysis is ongoing.

### ***Research Plans October 2006 through March 2007***

In October 2006 the results of usability testing conducted in summer 2006 will be reported at the metadata roundtable and submitted as GSLIS Technical Report.

We will continue conducting transaction log analysis of the queries submitted to the IMLS DCC Collection registry in 2006 in conjunction with the [same-procedure] analysis of the user queries in the Item Repository.

Recent online survey and 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.

We will also survey some of the LSTA grantees regarding this community's readiness to participate in the collection registry and item-level repository. In particular, we are interested if individuals involved in overseeing creation and management of LSTA projects/collections are emphasizing or including digital component in LSTA programs, and what is the best way to add collection information and item-level records (e.g., at state level or individual project level).

## **Related Activities**

### ***Metadata Roundtable***

We continued to hold the metadata roundtable study group with a weekly frequency. We are very pleased that the number of the regular participants in the metadata roundtable has increased. Recent roundtable topics have included *Is Metadata Dead?*; *The Framework of Guidance for Building Good Digital Collections*; Collection Definition; Calhoun Report; Subject Access to Federated Collections: a Case of IMLS DCC Collection Registry; and Whole-part Relationships and Boundaries in the Context of VRACore Metadata. 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>.

### ***Report on the Framework of Guidance for Building Good Digital Collections***

The project team produced a report on how *The Framework of Guidance for Building Good Digital Collections* is being used by the digital library community. Results of surveys one and two inform this report, and a discussion of the *Framework* was brought to the metadata roundtable. The final version of the report includes sixteen recommendations to NISO suggesting potential ways to improve the impact and/or utility of the *Framework*, and eight related research opportunities that IMLS may wish to consider.