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

Interim Performance Report 5
October 2004 – March 2005

Submitted by Timothy W. Cole, Principal Investigator, and
Sarah L. Shreeves, Project Coordinator
University of Illinois at Urbana Champaign
April 2005

Grant LG-02-02-0281
Interim Performance Report 5
1 October 2004 – 31 March 2005
Submitted by
Timothy W. Cole, Principal Investigator
Sarah L. Shreeves, Project Coordinator
April 2005

Summary

The IMLS Digital Collections and Content (DCC) Project continues to make progress on its objectives. Collection-level descriptions contained in our IMLS NLG Digital Collection Registry were released to the National Leadership Grant (NLG) projects for editing and approval in January 2005, and the Registry itself was made publicly accessible in February in conjunction with the IMLS Web-Wise Conference. Usability testing on the Registry has been conducted with both IMLS staff and with representative librarian populations, and changes based on these results are planned. We have continued to work with IMLS funded institutions to set up Open Archives Initiative (OAI) metadata provider services. As of March 2005, the IMLS item-level metadata repository contained approximately 200,000 DC records from 31 OAI-compliant NLG projects. The research team has continued to work through the interviews, focus group results, and surveys it has collected, and is preparing for a major survey in the fall of 2005.

General Project Activities

Project Staffing
The IMLS DCC project has seen two major staffing changes this period. Funding for John Lewis, the half time research programmer since December 2002, ended as planned in December 2004. Sarah Shreeves, the project coordinator since December 2002, has taken another position at the University of Illinois and ends her tenure as project coordinator April 30. A new project coordinator is now being sought. Professor Shreeves will continue her association with the project as a co-PI through the end of the grant.

Timeline
The project has continued to move forward and has generally met its targeted milestones as adjusted in previous interim reports. The Collection Registry was released for vetting in January 2005 (a month later than stated in the last interim report) and was released to the public at the Web Wise Conference in February 2005. Subject to approval by IMLS, the item-level metadata repository will be released to the public in late May 2005. Continued usability testing and subsequent changes to the interface will occur during the summer and early fall. Other major

project activities that will continue through the fall 2005 and early spring 2006 (subject to IMLS approving a no-cost extension) are the investigation of how to best to provide simultaneous and integrated access to both item-level and collection-level descriptive records and continued work with active NLG projects to set up and test OAI data providers. The research team has also revised their timeline based on assumed public availability of item-level metadata repository in May. Revised timelines for both sides of the project are included in Appendix One.

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.

<u>Publications</u>

Palmer, Carole L. and Ellen M. Knutson. 2004. Metadata practices and implications for federated collections. In Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology , Edited by Linda Schamber & Carol L. Barry. Medford, NJ: Information Today, Inc: 456-462.

Stvilia, Besiki, Les Gasser, Michael Twidale, Sarah L. Shreeves, and Timothy W. Cole. 2004. Metadata quality for federated collections. In Proceedings of ICIQ04 - 9th International Conference on Information Quality. Cambridge, MA : 111-125

<u>Presentations</u>

Besiki Stvilia presented a paper on metadata quality based on data from the IMLS DCC project titled "Metadata Quality for Federated Collections" at the 9th International Conference on Information Quality at MIT, Boston on November 6, 2004.

Tim Cole presented "The Open Archives Initiative Protocol for Metadata Harvesting and the IMLS Digital Collections & Content Project at the University of Illinois" at the Museum Computer Network 23rd Annual Conference, Minneapolis, MN, November 12, 2004.

Project team members presented at the ASIS&T annual meeting in Providence, RI, November 12-17 2004:

> Carole L. Palmer presented "Metadata Standards Adoption and Knowledge-Sharing" as part of a panel titled *Diffusion of Knowledge in the Field of Digital Library Development: How is the Field Shaped by Visionaries, Engineers, and Pragmatists?* on November 16.

> Carole L. Palmer and Ellen M. Knutson presented "Metadata practices and implications for federated collections" on November 17.

> Besiki Stvilia, Les Gasser, Michael B. Twidale had a poster titled "Metadata and Assuring Metadata Quality in Federated Digital Collections" November 14-15.

Sarah Shreeves presented "Integrating Access to Digital Content or OAI is Easy, Metadata is Hard" at the Visual Resources Association 23rd Annual Conference, Miami Beach, FL, March 6, 2005.

Sarah Shreeves presented "Search Interoperability, OAI, and Metadata" at two of the IMLS funded Illinois Digitization Institute's Track 3 Online and Hands-On Courses at the University of Illinois, Urbana, IL, October 4, 2004 and March 1, 2005.

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). 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, and Tim Cole is leading a sub-group on best practices for OAI data provider implementations. In addition, Sarah Shreeves has been appointed the chair of the Aquifer Metadata Working Group, an initiative of the Digital Library Federation.

Upcoming Publications and Presentations

Several members of the IMLS DCC team have collaborated on a paper titled "Is 'quality' metadata 'shareable' metadata? The implications of local metadata practices for federated collections" to be presented at the Association of College and Research Libraries 12th National Conference in Minneapolis, MN in April 2005. Ellen Knutson and Sarah Shreeves will present.

A paper drawing on our experience in the IMLS DCC project and titled "Using Collection Descriptions to Enhance an Aggregation of Harvested Item-Level Metadata" co-authored by Muriel Foulonneau, Tim Cole, Tom Habing, and Sarah Shreeves has been accepted into the 2005 Joint Conference on Digital Libraries (JCDL) in Denver, CO in June 2005. Tim Cole and Muriel Foulonneau will present.

Tim Cole with Simeon Warner from Cornell University will be presenting a tutorial on "OAI-PMH repositories: quality issues regarding metadata and protocol compliance" at the CERN Workshop on Innovations in Scholarly Communication (OAI4) in Geneva, Switzerland in October 2005.

Sarah Shreeves has written a chapter presenting the basics of the OAI Protocol in an upcoming publication from Libraries Unlimited. See: Shreeves, S.L. (In press). The basics of the Open Archives Initiative Protocol for Metadata Harvesting. Chapter in N. Courtney (Ed.) Technology for the Rest of Us: A Primer on Computer Technologies for the Low-Tech Librarian. Westport, CT: Libraries Unlimited.

Sarah Shreeves has co-authored an article on current developments for the OAI protocol for the Spring 2005 issue of *Library Trends*. See: Shreeves, S.L., Habing, T., Hagedorn, K., & Young, J. (2005). Current developments and future trends for the OAI Protocol for Metadata Harvesting. *Library Trends* 53(4):576-589.

Tim Cole and Sarah Shreeves are collaborating with colleagues at UIUC to write a book on the Open Archives Initiative which will be published by Libraries Unlimited.

Steering Committee Activity
The steering committee met on February 18, 2005 directly after the IMLS Web-Wise Conference in Washington, D.C. The main focus of the meeting was to help the project determine what items should be concentrated on during the last phase, and what were the potential research questions for follow on work. The minutes of the Steering Committee meeting are included in Appendix Two.

Collection Registry Metadata Schema and Service

Survey of IMLS NLG Projects

There has been no activity with the web-based Surveys 1 and 2 since the last interim performance report. We are planning a third round of the survey in June 2005 for the 2003 / 2004 National Leadership Grant recipients as well as non-respondents to the first two rounds of the survey.

Collection Description Metadata Schema

Tom Habing, a research programmer at the Grainger Engineering Library and Information Center at UIUC, developed XML schemas for the IMLS DCC Collection Description Metadata Schema. These are available at http://imlsdcc.grainger.uiuc.edu/registry/default.asp. Because the metadata schema is composed of several different types of entities and their relationships to each other, METS is a natural choice for bundling the descriptive metadata for each of the entities into a single file and expressing the structural relationship between these entities. In addition to the METS bundle, we also developed our custom XML schema for bundling these multiple entities into a single file. However, METS provides a much richer framework for bundling and expressing structure than our custom approach. Schema work has been shared with the METS community, and the IMLS DCC project collection-level description schema and registry has been selected as the starting point for an upgraded registry of DLF collections being implemented at the University of Illinois to support the DLF Aquifer project. In addition several other institutions and organizations have expressed interest in the XML schema for the collection description schema.

Sarah Shreeves has continued to be an active participant in the Dublin Core Collection Description (CD) Working Group and the NISO Metasearch Initiative Collection Description Task Group. She continues to work on a Usage Guide for the Dublin Core CD Application Profile, based in part on her experiences in the IMLS DCC project.

IMLS Digital Collections Registry

In January 2005, 100 National Leadership Grant projects were invited to edit and/or approve their collection descriptions through the interface at http://imlsdcc2.grainger.uiuc.edu/collregistry/ (password protected). As of March 31, 2005 approximately 50 collection descriptions have been changed or actively approved, i.e. the project staff went through the edit/entry forms and clicked the "finalize changes" button. Many more of the collection descriptions may have been tacitly approved, but there is no way for the IMLS DCC project to track this. Once the edit/entry forms have been edited and approved by NLG project staff, the IMLS DCC project coordinator reviews the changes made, makes changes as necessary, and then uploads the changes to the production database which is displayed to the public. The coordinator is documenting the time needed for the review process; currently most time is spent on authority work on institution names. This will become a larger issue as the collection registry grows. Currently there are approximately 518 institution names within the database.

The IMLS Digital Collections Registry (http://imlsdcc.grainger.uiuc.edu/registry/) was made public at the Web Wise Conference in Washington DC in February 2005. Within the collection registry there are currently 148 National Leadership Grant collections (including sub-

collections). In addition, 29 non-NLG digital collections were noted as associated collections, and 61 physical collections were noted as the source for the NLG digital collection.

A series of usability tests was conducted with IMLS staff members at the IMLS Web Wise Conference, as well as with librarians participating in the Illinois Digitization Institute on-campus training in March 2005. The results of the usability testing are being analyzed. The next six months of the project will focus on changes to the interface based on the testing and additional testing conducted with staff in local museums and historical societies.

Item-Level Metadata Repository

Assisting projects in implementing OAI-data provider services
The number of NLG projects with OAI data provider services increased 3% in the last six months. The breakdown of 1998-2004 NLG projects (136 total) in relation to OAI-data provider services is:

Category of 1998-2004 NLG Projects:	Number / % of NLG Projects:
Group 1 – Projects with OAI data provider sites for NLG content	31 (23%)
Group 2 – Projects whose institutions have an OAI implementation (not yet being used for NLG content) and projects that have explicitly expressed plans to add OAI functionality	41 (30%)
Group 3 – Projects who meet certain technical criteria – e.g. have item-level metadata and a maintained web site	25 (18%)
Group 4 – Projects with no item-level metadata, no interest in providing metadata via OAI, or whose grants were given up	20 (15%)
Unknown	19 (14%)
Total	136 (100%)

We are in communication with several NLG projects about planned implementation of OAI data providers, including Connecticut History Online (LL-90087 and LG-30-02-0256) and the Global Performing Arts Database at Cornell University (LG-30-02-0235). We continue also to offer testing services for OAI data provider services of NLG projects and provide feedback on metadata issues. In addition to working with NLG projects, we also offered significant feedback to the group at the Getty Research Institute developing the CCO-Lite, a metadata schema based on the CCO (Cataloging Cultural Objects).

We continue to contact NLG recipients about implementing OAI-data provider services and to document the barriers to implementation. Our goal by the end of the project remains to have approximately 50% of all NLG projects providing or committed to providing metadata via OAI. In the last phase of the project we will be concentrating on active projects as advised by the IMLS DCC Steering Committee. We also continue to document the obstacles to OAI data provider implementations and plan to issue a white paper examining these in Summer 2005.

Metadata harvesting and design of item-level repository
We have continued to harvest and index item-level metadata from NLG projects. As of March 31, 2005 we had harvested approximately 200,000 DC records from 31 OAI-compliant NLG projects. The repository is currently available at: http://imlsdcc.grainger.uiuc.edu/search/. Sites and number of records harvested as of March 31st, 2005 is available in Appendix Three.

We have slightly changed the interface to the metadata repository by adding a header and footer as well as a page documenting which collections are included. The next phase of the project will concentrate on interface development and usability testing of the item level metadata repository. In addition we plan to investigate how best to link the metadata repository and the collection registry so users can easily move between the two.

Research

Data Collection

Transcription of all 28 interviews is complete and analysis of the data is ongoing. We are developing coding schemes and working through intercoder reliability. We recently focused analysis on metadata quality for the paper the project team accepted for ACRL with efforts now turning toward users and repository applications. We are continuing to work with NLG projects to collect user studies, and are planning a new round of interviews that focuses on collection identity in the registry and perceptions of harvested metadata. We conducted a focus group at WebWise 2005 that focused on metadata functionality, interoperability, and changes in project goals over time. The focus group was comprised of seven participants representing six projects from various types of institutions including academic libraries, public libraries, museums, and historical societies.

Also at WebWise we began to conduct usability tests of the collection registry. Five individuals participated in this usability testing. Back at UIUC we have also done small scale usability studies with representative user populations. Initial findings focus on improving the collections registry, including aspects of internal navigation, improving geographic browsing, clarifying available collections metadata, and addressing variations in terminology.

Research Plans April 2005 through December 2005

Our research plan for year three is slightly modified from our original plan. We had to postpone the second round of phone interviews from the end of year two to spring of year three due to the delay in the roll out of the collection registry and item level repository. For the same reasons we postponed the customized user survey as well as the second survey to be sent to all NLG recipients with digital content. Unfortunately, in this cycle we will only be able to collect the data from the second survey, we will only have time for very limited, preliminary analysis of the data. We will continue with the usability testing with several different types of user groups. See Appendix One for the updated timeline.

Related Activities

We continue to have bi-monthly metadata roundtables. Recent roundtable topics have included the VRA Cataloging Cultural Objects and Learning Objects Metadata. This 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.

Appendix One
Revised Timelines

Collection Registry / Item Level Metadata Repository
(subject to approval of no-cost grant extension)

	Winter 05 Jan-Mar	Spring 05 Apr-Jun	Summer 05 Jul-Sep	Fall 05 Oct-Dec	Winter 06 Jan-Mar	Spring 06 Apr-Jun
Item Level Metadata Repository		♦ (public release) changes	Usability testing / iterative design			
		Metadata analysis / normalization / clustering				
Collection Registry		♦ (public release)	Usability testing / iterative design changes			
Merger of collection registry / item level metadata repository			Design / Development	♦ (public release) Usability testing		
Build / test OAI data providers in active projects	Continued contact / with NLG projects about OAI / interoperability / metadata					
Survey of NLG projects			♦ 2003/04 and non respondents			
Reports / White Papers			August 2005 – Obstacles to OAI and interop.	October 2005 – Interim Performance Report		May 2005 – Final Report

Research Team
Year Three 2005

	Winter Jan-Mar	Spring Apr-Jun	Summer Jul-Sep	Fall Oct-Dec
Focus Group at WebWise (n=7)	Metadata, Users, Interoperability			
Phone Interviews (n~15-20)		Users, Repository Applications, Interoperability		
Customized User Survey (n~10)			Communities, Repository Applications	
Survey 2 (n=136)				Metadata, Repository Applications, Interoperability
Use and Usability Testing (n=8)		Interoperability		

Appendix Two
Notes from the IMLS DCC Steering Committee Meeting

IMLS DCC Steering Committee Meeting
February 18, 2005 -- 12:30-2:30
Hyatt Regency on Capitol Hill, Washington, DC

Attending:
Liz Bishoff, Tim Cole, Anne Craig, Martha Crawley, Jane Greenberg (by phone), Nuala Koetter, Carole Palmer, Sarah Shreeves (facilitating), Brenda Trofanenko, Mike Twidale

Action items:
- Disseminate research findings on web site as much as possible
- Investigate organizing a steering committee meeting at ALA

Priority Areas

Collection Registry:

Immediate items:
- Understand how NLG projects are completing/modifying their collection description records
- Determine what pieces of the review of records in the registry we can automate (through validation, etc)
- Integrating the collection and item level registry

Continued and future research topics:
- What is a collection in the digital environment?
- How are the different communities (libraries, museums, archives, etc) thinking about collection identity? Are these coming closer together in the digital world?
- When is the concept of collection useful in the federated environment? How do we effectively integrate collection and item level descriptions in a federated environment?

Item Level Metadata

Immediate items:
- Prepare item level metadata repository for public release
- Identify and report on best practices for shareable metadata specifically for IMLS

Continued and future research items:
- There is a need for tools and services to enhance, enrich, and modify metadata in automated ways.
- How can we help the different cultures of description create better quality metadata?
- Is there a place for open source cataloging? For allowing a group of users to catalog items? What are the authority issues here?

- How can we build in the idea of creating 'shareable' metadata from the very start of the digitization project? To transition it from an add-on at the end to something that is consciously done from the very start?

OAI uptake

Immediate items:
- Work with the projects that are still active to establish OAI data providers.
- Encourage CONTENTdm projects to upgrade to the latest version of the product to take advantage of using Qualified Dublin Core.
- A report to IMLS on the resources needed to implement OAI data provider services as well as the the barriers to implementing OAI.
- Continue work on the OAI best practices work, but develop some guidelines specific to the type of institutions funded by IMLS (perhaps in conjunction with Emory?).

Continued and future research items:
- How can we build in the idea of interoperability from the very start of the digitization project? To transition it from an add-on at the end to something that is consciously done from the very start? [Related to item above]

Other Issues and Questions

- Yes, it would be useful to include more public libraries (than is available through the NLG projects) in the research portion. Possible avenues are IDA or other state funded programs. PLA would be an important dissemination point.
- Knowledge diffusion is very important particularly among collaborations. We also need to understand where knowledge diffusion worked and where it did not.
- What does all this mean for the next round of IMLS funding? Perhaps planning a steering committee meeting at ALA to discuss the implications of the project's work for IMLS, its grantees and others would be useful.

Appendix Three
National Leadership Grant Collections and Number of Records Harvested

189,334 DC records from 31OAI-compliant NLG projects as of March 31, 2005

Academy of Natural Sciences
"American Natural Science in the First Half of the Nineteenth Century" - LL-90013
349 records

Alliance Library System
"Illinois Alive!" - LL-80052
112 records

Brandeis University
"Honore Daumier Lithographs" - ND-10005
3,873 records

Broward County Library
"Education by Design: Educational visual aids from the Bienes Center's WPA Museum Extension Project Collection" - ND-10010
666 records

Colorado Digitization Program
"Heritage Colorado" - LL-90094
25,629 records

Colorado Digitization Program
"Western Trails" - NL-10024
6,848 records

Cornell University
"Kinetic Models for Design Digital Library (KMODDL)" - LG-30-04-0204
480 records

Greater Western Library Alliance
"Western Waters Digital Library" - LG-03-03-0113
28,003 records

Indiana University
"Charles W. Cushman Photograph Collection" - ND-00022
14,425 records

Louisiana State University
"Louisiana Purchase Bicentennial: A Heritage Explored" - ND-00010
714 records

Michigan State University
"Feeding America: The Historic American Cookbook Project" - ND-10042

77 records

MOAC
"MOAC: Collections in California Museums" - LL-90130
30,843 records*

Tufts University
"Bolles Archive of London" - ND-00015
35 records

Tulane University - Amistad Research Center
"American Missionary Association and the Promise of a Multi-cultural America:1839-1954" - LL-90044
3,342 records

University of California, Riverside
"INFOMINE Scholarly Internet Resource Collection" - LG-02-03-0083
81 records

University of Central Florida
"Central Florida Memory" - LG-30-04-0225
66 records

University of Florida. Florida Center for Library Automation.
"Florida Environmental Information Online" (Part of "Linking Florida's Natural Heritage" - LL-80016)
1,155 records

University of Georgia / University of Tennessee
"Southeastern Native American Documents" - LL-90019 and ND-00017
289 records

University of Illinois
"Teaching with Digital Content" - NL-00003
1,558 records

University of Maine
"Maine Music Box" - LG-03-02-0116
13,261 records

University of Michigan
"Flora and Fauna of the Great Lakes" - NL-00034
26,300 records

University of Minnesota
"Summons to Comradeship: World War I and II Posters" - ND-10007
5,740 records

University of North Carolina
"Southern Homefront" - LL-80202
405 records

University of North Carolina
"North Carolina in Black and White" - ND-00031
439 records

University of Tennessee
"Tennessee Documentary History" - ND-10020
1,314 records

University of Tennessee
"Frank H. McClung Museum WPA/TVA Photograph Archive" - LG-03-02-0080
6,317 records

University of Washington / Museum of History and Industry
"King County Snapshots: A Photographic Heritage of Seattle and Surrounding Communities" -
NL-10016
12, 213 records

University of Wisconsin-Madison
"Africa Focus" - LL-80131
3,847 records

Washington State University
"Columbia River Basin Ethnic History" - NL-10032
774 records

Wisconsin Historical Society
"American Journeys: Eyewitness Accounts of Early American Exploration and Settlement" -
LG-03-02-0112
179 records

* We are experiencing harvesting problems with MOAC.