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

Interim Performance Report 1
1 October 2002 – 31 March 2003

Submitted by Timothy W. Cole, Principal Investigator, April 2003
University of Illinois at Urbana Champaign

Institute of Museum and Library Services

Digital Collections and Content

Working toward interoperable digital content.

Grant LG-02-02-0281
Interim Performance Report 1
1 October 2002 – 31 March 2003

Submitted by Timothy W. Cole, Principal Investigator, April 2003

Summary

The first six months of the University of Illinois IMLS Digital Collections and Content project were spent refining project planning, hiring staff, purchasing and configuring project servers and workstations, forming a project steering committee, creating a project Web site, and developing a project evaluation plan. Other activities included reviewing information about more than 90 relevant National Leadership Grant (NLG) projects, creating documents for prioritizing and tracking interactions with grantees, disseminating information to grantees, creating survey instruments, and implementing a simple item-level metadata repository for preliminary testing. We also reviewed community collection metadata schemes, drafted a project-specific collection metadata scheme, and refined and began work on our three-year research plan. Implementation of services has been slow as compared to proposal schedule, but significant progress has been made and we anticipate catching up in this regard before next report.

Start-Up Activities

Hiring staff, configuring equipment, and creating Web sites

Three positions were advertised and filled.
- Sarah Shreeves was hired as project coordinator. (Full time; 12/03/02)
- John Lewis was hired as research programmer. (Half time; 12/09/02)
- Ellen Knutson was hired as project research assistant. (20 hrs/wk; 01/21/03)

Necessary equipment was purchased and deployed.
- A workstation was purchased for the project coordinator.
- Disk, server, and application (e.g., database) workspace assigned and configured.

Project Web sites and a steering committee listserv were established.
- Project: http://imlsdcc.grainger.uiuc.edu.
- Steering committee: http://imlsdcc.grainger.uiuc.edu/steeringcommittee/ (password required).
- Steering committee listserv: imls-steercom-l@listserv.uiuc.edu (closed list)

Forming a steering committee

In consultation with IMLS, an eight-member steering committee was formed in December 2002. Steering Committee members are Liz Bishoff (Colorado Digitization Project), Priscilla Caplan (Florida Center for Library Automation), Anne Craig (Library Automation & Technology, Illinois State Library), Martha Crawley (*Liaison*, Office of Library Services, IMLS), Jane Greenberg (University of North Carolina at Chapel Hill), Richard Rinehart (Berkeley Art Museum), John Saylor (Cornell University), and Brenda Trofanenko (UIUC). The first annual meeting of the steering committee was held in February 2003 in conjunction with the WebWise Conference in Washington, D.C. (Meeting report attached in Appendix 1.)

Attending conferences and training

In Dec. 2002, Tim Cole and Sarah Shreeves attended the IMLS-sponsored Outcomes-Based Evaluation Training in Washington, D.C. As a result of this training, a logic model was created to identify immediate, intermediate, and long-term outcomes and measures. (Model attached in Appendix 2.) In January 2003, Sarah Shreeves attended the Open Forum on Metadata Registries in Santa Fe, NM. The conference provided useful context for the IMLS DCC project. In February 2003, Sarah, Tim, and Ellen Knutson attended WebWise in Washington, D.C. IMLS DCC was promoted via a printed brochure and contacts with NLG projects were made. In March 2003, Mike Twidale attended Museums and the Web in Charlotte, NC.

Collection Registry Metadata Schema and Service

Designing survey of IMLS grantees

In our grant proposal, we stated that a survey of NLG projects would be conducted within the first six months of the project to allow us to ascertain the nature and scope of extent NLG collections, assess the volume and availability of item-level metadata, and identify availability of ancillary content like project transaction logs and user studies. In December 2002, Martha Crawley of IMLS furnished us with copies of the original proposals for all NLG projects funded between 1998 and 2002 with associated digital content. Proposals and Web sites for each project were examined, and 98 projects were identified as likely participants. We created a spreadsheet to track NLG project details, facilitate contacts, and support our research. The spreadsheet includes contact details, grant details (descriptions and collaborators), content details (topic areas and metadata schemas), and technical details (databases in use and OAI protocol compliance) as best as can be determined from examination of Web sites and proposals. The spreadsheet was completed in early February and has since been shared with IMLS.

As anticipated, NLG project staff, responsibilities, and focus often shift during the life of a project and after its completion. As a result, details in grant proposals and on project Web sites are often out of date. Some needed information about projects and their context can not be deduced from available information. A survey is still required; however the information gathered through inspection of proposals and review of Web sites was very helpful in developing survey instruments. Also helpful was input from the Steering Committee. While these processes and review delayed submittal of survey instruments and plans to the Office of Management and Budget (OMB), our final survey plan is considerably more sophisticated than originally envisioned, and will therefore yield better results.

Two survey instruments designed to be mailed together to NLG grantees with projects having associated digital content were created. The first instrument will be used to collect and verify project and collection information so that a preliminary registry can be created and so that

technical readiness of NLG projects to provide item-level metadata can be assessed. The second instrument will support our research investigations into the use and usefulness of metadata by soliciting baseline information about projects' experiences with metadata creation and their opinions on the value and purpose of a collection registry. We will follow up mailing of the surveys with email and phone calls. On March 21, 2003, the University's Institutional Research Board Office assessed the IMLS surveys as having no significant risk for participants. The Office also waived review by the full board. On March 23, 2003, a copy of the survey was provided to IMLS for formal submittal to OMB. Assuming OMB approval by early June, we will distribute the survey instruments in June and analyze results by mid-3rd quarter of 2003.

Defining collection-level metadata schema

To help us define an IMLS DCC metadata schema, we conducted a literature review and a search for available collection-level metadata schemas. We identified three of interest:

- A collection description (CD) schema developed for the Research Support Library Programme (RSLP) in the UK.[1]
- A proposed Dublin Core (DC) collection description schema, based on RSLP.[2]
- The Encoded Archival Description, used for archival finding aids.[3]

In January, we held a conference call with Pete Johnston, co-chair of the DC Collection Description Working Group and interoperability research officer in the United Kingdom Office for Library Networking (UKLON). We discussed implementations of and plans for the RSLP CD and DC schemas. Pete later shared with us latest documents for expressing the RSLP schema in RDF/XML and XML. In early February, we presented the three schemes listed above to the project steering committee. The steering committee reached consensus that the registry should focus on collections, not projects, and it should allow for multiple collections covered by large projects. NLG projects may be transient, but the digital collections they create will (hopefully) have long lives. We also explored at length the related issue of how metadata is inherited from projects and collections. The steering committee favored more complex collection-level descriptions (e.g., RSLP) and warned us to avoid over-simplified schemas. During February and March 2003, a customized collection metadata scheme for this project was developed based on the RSLP scheme and with explicit mappings to the DC collection scheme. The scheme includes selective use of standard vocabularies and formats. Scheme details were forwarded to IMLS and the steering committee in early April 2003.

Designing and building the collection registry

Several existing, active collection registries were examined for functionality, interface design, and metadata schema. They included NSDL (http://www.nsdl.org) and Cornucopia (http://www.cornucopia.org.uk). In January, we consulted with David Dawson of Resource: The

---

[1] See http://www.ukoln.ac.uk/metadata/rslp/schema/
[2] See http://homes.ukoln.ac.uk/~lispj/dc-cd/rslpcd.html
[3] See http://www.loc.gov/ead/

Council for Museums, Archives & Libraries in the UK about Cornucopia and plans to develop a registry for the NOF-Digitise project (UK) and the Minerva project (Europe). Through

examination of these registries and conversations with David, Pete, and others, we have identified several functions to be included in the IMLS DCC registry. They include browsing by topic area, expressing relationships among collections (parent-child, etc.), and limiting searches by time period, geographic area, audience, or type of material.

During the next six months, we will build the database structure for the registry; design, develop, and test the forms that will enable NLG projects to enter and maintain collection metadata; and design, develop, and test the registry interface. We expect to complete an alpha mock-up of the registry by the end of the 2nd quarter of 2003, and we expect a beta version by the end of the 3rd quarter. New target for completing a production version of the registry is December 2003.

Item-Level Metadata Repository

Analyzing item-level metadata and schemas

As a result of our preliminary analysis of NLG grant proposals and projects, we have a better understanding of the range of item-level metadata schemas in use. This understanding will be augmented by the survey results. We have documented the use of the following schemas: MARC, DC (simple and qualified), TEI, TEI Lite, EAD, Federal Geographic Data Committee (FGDC) standards, VRA Core, and locally developed schemas.

OAI mandates the use of DC, but also supports optional use of other, schemas. We have developed a metadata schema, called IMLS_DC, for possible use in our item-level repository. It is based on qualified DC and work done by both the NSDL and the Western States Digital Standards Group.[4] Our experience on a previous project, the Illinois OAI Metadata Harvesting Project (funded by the Andrew W. Mellon Foundation), shows that although use of DC allows a reasonable level of interoperability, the use of qualifiers is useful for reflecting the complexity of aggregated metadata that describes many resources. Greater levels of interoperability can be achieved through the use of qualified DC. As we identify additional elements needed by NLG projects and IMLS, we will continue to update and enhance the IMLS_DC schema.

Assisting projects in implementing OAI-metadata provider services

We have preliminarily segmented NLG projects into four groups, from three of which we will solicit involvement in the item-level repository. These groups have been prioritized as follows:

- **Group 1:** Projects with OAI data provider sites for NLG content.
- **Group 2:** Projects whose institutions have an OAI implementation (not yet being used for NLG content) and NLG projects that have express plans to add OAI functionality.
- **Group 3:** Projects who express enthusiasm and meet certain technical criteria (e.g. have item-level metadata and a maintained web site).
- **Group 4:** Projects with no item-level metadata or no interest in providing metadata via OAI.

---

[4] See http://www.cdpheritage.org/westerntrails/wt_bpmetadata.html.

As indicated in our proposal we will provide direct assistance (remotely) to projects wanting help implementing OAI. In February 2003, we completed remote installation of an OAI data provider

service for the Colorado Digitization Program (CDP). This service was implemented on top of the metadata storage infrastructure already in use by CDP and did not require changes to metadata processing or workflow. The CDP service supports exporting metadata in the IMLS_DC schema mentioned above, as well as in simple DC. (Generic versions of the all University of Illinois Library metadata provider implementations and associated XML schema definitions created as part of this work are available on SourceForge under UIUC/NCSA Open Source licensing.) Initial conversations with several NLG projects about setting up OAI-metadata provider services at their sites are also underway.

Implementing metadata harvesting and repository

In February, we implemented an alpha item-level repository using our OAI harvester. We harvested approximately 20,000 DC records from seven OAI-compliant NLG projects and stored them in a SQL database. The bulk of the records were harvested from CDP. The repository is available at http://rama.grainger.uiuc.edu/searchimls/ (password protected). By the end of the 3rd quarter, we will complete a beta version of the repository, harvesting metadata as available from OAI-compliant NLG projects.

Research

The research subgroup for this project has met on a weekly basis since the end of January. Initial meetings focused on creating a three-year research plan and the survey instruments discussed above. In creating the research plan we decided to focus initially on a series of sub-questions and data collection methods that would help us answer the primary research question:

> How can resource developers best represent collections and items to meet the needs of service providers and end users?

Research plan identifies six sub-topics for investigation: baseline information, metadata, users, single collection use by community, repository applications, and interoperability. Each of the sub-topics has several data collection methods associated with it, and most will be analyzed across time. We plan to use several survey instruments over the course of the three year research projects, as well as interviews and focus groups. (Research plan attached as Appendix 3.)
In addition to identifying and reading background materials on metadata, the research group reviewed the 98 NLG proposals described above and identified prospective projects for targeted research. In order to assure a diversity of sites we used several variables to select prospective targeted sites including size and type of institution, size and type of collection, choice of metadata scheme, and whether or not it was a collaborative project. As mentioned above, initial contact was made with a few of these potential research participants at WebWise.

In order to best learn from other ongoing research the group also has identified related research at the Graduate School of Library and Information Science. We are planning regular meetings with members of these research projects to discuss our common interests in metadata. Further, we helped draft a UIUC Research Board proposal for a related project on metadata quality. In

the next six months the research group will make formal initial contact with the 15-20 targeted research sites and begin to set up phone interviews. We will continue with the content analysis of the project proposals and analysis of survey data as it is returned.

Outcomes Logic Model

Organization Name:	University of Illinois at Urbana-Champaign		
Project Name:	IMLS Digital Collections and Content		
Date Created	12-6-02	Date Reviewed	

Program Influencers *(Key entities that help define the program or to whom the program will report results)*
IMLS / National Leadership Grant (NLG) Grantees, Digital Library Community, Library School Faculty and Community, End Users

Organizational Mission *(Organization's mission statement or key action words)*

Program Purpose	
We do what? *(Summary of key proposed services)*	• Build collection registry • Build metadata repository • Investigate and prove benefits (research use and usability) • Investigate models for sustainability
For whom? *Target population(s)*	• IMLS • NLG Grantees • Other Digital Libraries (NSDL) • End Users • LIS Faculty
For what outcome(s)? *(Benefits/changes in skills, knowledge, attitude or life condition.)*	• Grantees have a new skill set (metadata creation and sharing) • Increased use and usability of IMLS NLG digitized content • Metadata sharing/interoperability incorporated into the LIS classroom • Library and museum community does better job including interoperability in digital library /collection designs

Inputs *(List items dedicated to or consumed by the program)*	Outputs *(Program products)*
1.5 FTE 2.25 GA 10% of 1 Faculty 5% of 4 Faculty 3% of 1 Faculty 7 Steering Committee members Facilities (Grainger Library computing infrastructure) Travel funding Metadata from NLG grantees Access to students for testing	# of grantees participating # of records in repository # of end users involved in usability testing # of transactions in use of registry # of presentations / reports / papers Metadata schemas developed Registry / Repository developed

Program Activities *(List key activities needed to provide or manage services.)*	**Program Services** *(List services to be delivered directly to participants.)*
Recruit NLG participants Research metadata schemas Coordinate project staff/ budget Design, develop, test, evaluate registry and repository Prepare reports / papers / presentations Evaluate outcomes	Provide metadata schemas Provide registry / metadata search interface Assistance / guidance to NLG grantees Provide reports to IMLS Report out to LIS community

Target Population *(List specific characteristics of primary intended participants)*

Intended Outcomes *(Changes in skill, knowledge, attitude, behavior, life condition or status)*	**Indicators (Measures)** *(Concrete evidence, occurrence, or characteristic that will show the desired change occurred)*
Intermediate: 1. Grantees know how to create good metadata for interoperability and how to provide access to their metadata	# and % of grantees who are able to correctly provide metadata via OAI or OAI static protocol AND # and % of providers contributing records that contain at least title, creator, date, type, identifier, and description or subject used according to Dublin Core Standards
Intermediate: 2. End users can more easily find IMLS funded digitized content	# and % of end users who can accomplish __ % of tasks in usability test in set time
Long Term: 3. Grantees will collaborate on digitization projects	# and % of grantees reporting collaboration as a result of registry / repository AND # of IMLS grants awarded including collaborations between different partners
Long-term: 4. NSDL and other aggregators can harvest metadata from IMLS NLG grantees	# and % of aggregators who harvest at least X collections AND # and % of NLG collections harvested at

	least once
Immediate to Long Term: 5. LIS education community perceives value in metadata sharing and updates curriculum to reflect this. (Immediate change in attitude and long term change in field)	# of publications concerning interoperability accepted AND # of LIS programs incorporating interoperability issues into curriculum
Immediate to Long Term: 6. Library and museum communities perceive value of creating metadata for interoperability. (Immediate change in attitude and long term change in digital library environment)	# and % of NLG grantees reporting value in creating metadata for interoperability AND # of project's presentations / publications concerning interoperability accepted AND # of IMLS grants awarded including interoperability AND # and % increase of registered collections that can be harvested by the OAI protocol

Outcome #1 <u>Grantees know how to create good metadata for interoperability and how to provide access to their metadata</u>

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
# and % of grantees who are able to correctly provide metadata via OAI or OAI static protocol	Harvesting service records	NLG Grantees	Quarterly	
# and % of providers contributing records that contain at least title, creator, date, type, identifier, and description or subject used according to Dublin Core Standards	Analysis of metadata records	Participating NLG grantees	End of project and as each comes online	

Outcome #2 <u>End users can more easily find IMLS funded digitized content</u>

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
# and % of end users who can accomplish __ % of tasks in usability test in set time	Usability tests	Specific groups of users participating in testing	1 series in '03-'04 More in '04-'05	

Outcome #3 <u>Grantees will collaborate on digitization projects</u>

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
# and % of grantees	Survey	NLG Grantees	End of project	

reporting collaboration as a result of registry / repository				
# of IMLS grants awarded including collaborations between different partners	IMLS awarded grants	IMLS awarded grantees	Each award cycle	

Outcome #4 NSDL and other aggregators can harvest metadata from IMLS NLG grantees

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
# and % of aggregators who harvest at least X collections	Survey	Aggregators	End of project	
# and % of NLG collections harvested at least once	Provider side transaction records	Participating NLG Grantees	Annually	

Outcome #5 LIS education community perceives value in metadata sharing and updates curriculum to reflect this. (Immediate change in attitude and long term change in field)

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
# of publications concerning interoperability accepted	Literature survey	LIS faculty	1 midpoint and 1 end of project	
# of LIS programs incorporating interoperability issues into curriculum	Survey	LIS faculty and administrators	Follow on (after project)	

Outcome #6 Library and museum communities perceive value of creating metadata for interoperability. (Immediate change in attitude and long term change in digital library environment)

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
# and % of NLG grantees reporting value in creating metadata for interoperability	Survey	Participating NLG grantees	End of project	
# of project's presentations / publications concerning interoperability accepted	Literature survey	Project staff	Annual	
# of IMLS grants awarded including interoperability	IMLS	IMLS Grantees awarded	Each grant cycle	
# and % increase of registered collections that can be harvested by the OAI protocol	OAI site records	OAI registered providers	Midpoint and end of cycle	

IMLS DCC
Research Plan
Carole Palmer, Michael Twidale, Ellen Knutson
14 February 2003

Primary research question

> How can resource developers best represent collections and items to meet the needs of service providers and end users?

Specific research questions and methods of data collection

1) **BASELINE:** What are the different kinds of institutions, collections, and metadata schemes and interoperability represented in the IMLS projects? How do the resource developers anticipate their digital collections will be used?

2) **METADATA:** What metadata scheme(s) is being applied for item description?
   A. Why was the scheme selected?
   B. What are its strengths and weaknesses?
   C. How well does it (seem to) support various functions such as, administration, authentication, discovery, use?
   D. How well does it support the inevitable evolution of needs, technologies and opportunities which the organization wishes to exploit?

3) **USERS:** Who are the end user communities for these institutions (known and potential)?
   A. What are the institutions' perceptions of the value and level of use of the collections by these communities?
   B. How do those perceptions change over time?

4) **SINGLE COLLECTION USE BY COMMUNITIES:** How well do the collections and their metadata support the needs and practices of these communities? Can they search for and find materials in ways that make sense to them?

5) **REPOSITORY APPLICATIONS:** How do institutions hope to provide aggregate collections or services for their users? Does the registry and repository offer adequate support? Can they locate and aggregate materials that are meaningful for their constituencies? What are additional applications for the repository?

6) **INTEROPERABILITY:** What are the challenges in supporting interoperability?
   A. What are the problems in harmonizing metadata standards?
   B. What are the design challenges in developing interoperability technologies?
   C. What are the usability challenges for systems supporting interoperability?

Timeline and Methods

Year One 2003

	Winter Jan-Mar	Spring Apr-Jun	Summer Jul-Sep	Fall Oct-Dec
Content Analysis (n=94)	Baseline, Metadata, Users			
Survey 1 (n=94)			Baseline, Metadata, Users, Repository Applications, Interoperability	
Email follow-up (n=94)			Metadata	
Phone Interviews (n~15-20)			Metadata, Users, Repository Applications, Interoperability	

Year Two 2004

	Winter Jan-Mar	Spring Apr-Jun	Summer Jul-Sep	Fall Oct-Dec
Focus Groups at WebWise (n~8-20)	Metadata, Users, Interoperability			
Log Data (n=?)		Users, Communities		
Customized User Survey (n~10)			Communities, Repository Applications	
Phone Interviews (n~15-20)			Metadata, Users, Repository Applications, Interoperability	

Year Three 2005

	Winter Jan-Mar	Spring Apr-Jun	Summer Jul-Sep	Fall Oct-Dec
Focus Groups at WebWise or JCDL (n~8-20)	Metadata, Users, Interoperability			
Survey 2 (n=94)		Metadata, Repository Applications, Interoperability		
Phone Follow-up (n=?)			Metadata, Repository Applications	
Use and Usability Testing (n=?)		Interoperability		