

IMLS Digital Collections and Content

Grant LG-02-02-0281



Interim Performance Report 9 1 October 2006 – 31 March 2007

Submitted by:

*Timothy W. Cole, Principal Investigator
(217) 244-7837
t-cole3@uiuc.edu*

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

*April 2007
University of Illinois at Urbana Champaign*

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 9
October 2006-March 2007

Submitted by
Timothy W. Cole, Principal Investigator
Amy Jackson, Project Coordinator
April 2007

Summary

The IMLS Digital Collections and Content project continued to make progress during the performance period toward proposal goals and outcomes. New collections were added to both the registry and repository, data was migrated to a new server, and exploration of harvested metadata enrichment processes continued. A new interface for the item-level repository was tested and will be incorporated into the main site during the next performance period. The 2005 NLG projects have been contacted and records are currently being added to the registry. The Illinois State Library provided information for LSTA projects, and the registry and repository are currently being populated with these records. Additionally, the final Steering Committee meeting was held during the 2007 WebWise conference and feedback and recommendations were solicited from the members.

General Project Activities

Registry

The Illinois State Library provided information regarding all LSTA funded digitization projects in the state of Illinois. Collections hosted by the Illinois Digital Archives were identified as our next set of LSTA collection targets, and contact was made with projects managing 14 of these collections. Seven of the collections completed survey one and initial password protected records were created in the registry. Five of these collections approved these records and the collection records are now publicly available. Follow up with non-responding collections will occur in the next performance period.

Seventeen 2005 NLG projects were contacted with requests to complete survey one. Seven of these projects completed survey one, and initial password protected records were completed for these projects. Four of these projects have responded to date, and these records are publicly available. Follow up with non-responding projects will occur in the next performance period.

At the end of the current performance period the collection registry contained 178 collections contributed by 135 projects. This included 10 LSTA collections.

Repository

All new and previously added collections were examined in this performance period for ability to contribute metadata through OAI-PMH. Five collections added to the registry in previous performance periods were harvested, resulting in 46,137 new metadata records. Five LSTA collections in the Illinois Digital Archives were harvested, providing 727 new records. Appropriate 2005 NLG projects were harvested, and 1,350 new records have been added to date from these collections. At the end of this performance period the metadata repository contained 304,000 records from 37 collections.

Hardware

All IMLS DCC databases were migrated to a new server in December for enhanced performance and to facilitate anticipated implementation of new search interface.

Website and Search Developments

Preliminary analysis of transaction logs for the IMLS DCC website was undertaken. See Appendix One for results. Further log analysis is planned in the next performance period.

A new item-level interface underwent additional testing and refinement during this period, and harvests were refreshed to populate the underlying database. Additionally, item-level metadata reprocessing and augmentation were integrated into the regular workflow. Augmentation of metadata records includes addition of collection-level information. Public release of the new interface should occur in early April.

Timeline

Previously determined goals and developments progress according to the project timeline. In this performance period transactions logs were analyzed, post-harvest normalization of all content was implemented, processing and maintenance were tested and refined and 2005 NLG projects and select LSTA collections were added to the registry.

Dissemination

Presentations and Publications

In October 2006, preliminary results of the usability testing of the most recent iteration of the Registry search interface with a focus on two important but distinct constituencies of potential users – librarians and teachers – were presented at the IMLS DCC metadata roundtable. The PowerPoint presentation is available at https://netfiles.uiuc.edu/kgroets2/www/IMLS_Usability.ppt.

In November 2006, we presented the paper focusing on collection identity – *Palmer, C., Knutson, E., Twidale, M., & Zavalina, O. (2006). Collection Definition in Federated Digital Resource Development* – at the American Society for Information Science and Technology annual meeting in Austin, Texas. The paper has been published in Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology.

In early 2007, the chapter focusing on metadata quality in Digital Collections and Content Registry has been published -- *Stvilia, B., Gasser, L., Twidale, M. (2007). Metadata quality problems in federated collections. In L. Al-Hakim (Ed.), Information quality management: theory and applications. Hershey, PA: Idea Group.* Another paper has been submitted for review -- *Stvilia, B., Gasser, L. (2006). Value based metadata quality assessment.*

Sarah Shreeves presented at the 2007 WebWise Pre-Conference session, *Sharing Images and Data: Making Access to Collections Easier and Better.* Washington, DC, February 28, 2007.

In March 2007, the presentation *VRA Matchmaking Lessons Learned from the IMLS Digital Collections and Content Project* was delivered as part of the From Fair Use to Fair Trading: Creating a Digital Image Matchmaking Commons panel at the 25th Annual Conference of the Visual Resources Association in Kansas City, Missouri.

Upcoming presentations

The presentation entitled *Investing in Collection Representation for More Useful Repositories*, which is building on results from IMLS DCC project, will be delivered by Carole Palmer as part of the Revisiting the Foundations of Information Discovery and Access panel at the 70th Annual Meeting of the American Society for Information Science and Technology (ASIST) in October 2007 in Milwaukee, Wisconsin.

In October 2007, Michael Twidale will be leading The Live Usability Lab: Open Access Archives and Digital Repositories panel discussion at the 70th Annual Meeting of the American Society for Information Science and Technology (ASIST) in Milwaukee, Wisconsin.

Our paper – *Palmer, C., Zavalina, O., & Mustafoff, M. (2007). Trends in Metadata Practices: A Longitudinal Study of Collection Federation* – submitted to 2007 Joint Conference on Digital Libraries to be held in Vancouver, British Columbia, in June 2007, has been accepted and will be published in conference proceedings.

A paper reporting results of the transaction log analysis with the focus on types of searches and semantic similarities between user terms and controlled vocabulary terms – *Zavalina, O. (2007). Collection-Level User Searches in Federated Digital Resource Environment* – has been submitted to the 70th Annual Meeting of the American Society for Information Science and Technology (ASIST) to be held in October 2007 in Milwaukee, Wisconsin.

Lessons learned from the IMLS DCC project will be presented by Amy Jackson, Kurt Groetsch, Oksana Zavalina and Timothy Cole at the 2007 LITA Forum. *Aggregating collections for access: the IMLS Digital Collections and Content project at the University of Illinois at UC*. 2007 LITA National Forum, Denver, Colorado, October 5-7, 2007.

Other Forums

Sarah Shreeves, building off of the research and experience of the IMLS DCC project, has been actively presenting and conducting workshops on the OAI PMH and shareable metadata. These include workshops for the Western New York Library Resources Council and the METRO Library Consortium in NYC as well as a presentation at WebWise 2007. In addition, Sarah, Tim Cole, and Jenn Riley from Indiana University received a two-year Laura Bush 21st Librarianship Program grant from IMLS to develop an on-line and on-site training program for shareable metadata.

Steering Committee

The Steering Committee met on March 1 at the 2007 WebWise conference held in Washington, DC. The project team solicited feedback and advice on findings from the project, including metadata trends, usage log analysis, metadata for educational use, and collection identity. See Appendix Three for Steering Committee meeting notes.

Research

Metadata harvesting

Metadata records harvested through the course of the project were analyzed to determine trends in Dublin Core usage and mapping from the inception of OAI-PMH. Results to date are inconclusive. Analysis will continue during the next performance period.

Data collection and analysis

Quantitative and qualitative data collected through the 2nd round online survey, conducted in Spring 2006 with the focus on tracing material- metadata- and audience-related changes over time, have been analyzed and compared with the previous findings (from the 1st round of survey conducted in 2003-2004). Detailed comparative analysis is presented in a full paper accepted to JCDL'07 conference: *Palmer, C., Zavalina, O., & Mustafoff, M. (2007). Trends in Metadata Practices: A Longitudinal Study of Collection Federation*. This paper also incorporates analysis of qualitative data obtained through interviews with current and recent NLG recipients at the Web Wise'2006, focusing on challenges and successes in metadata selection and application. Results of our analysis demonstrate that although changes in metadata applications have not been pronounced, multi-scheme use has become less common, and use of Dublin Core remains high, even as recognition of its limitations grows. Locally developed schemes are used as much as MARC, and may be on the increase as new collections are incorporating less traditional library and museum materials,

and more interactive and multimedia content which is not easily described with available standards¹. Additional analysis based on paper reviewers' suggestions is ongoing.

One follow-up interview has been conducted at the Web Wise'2007 conference in March 2007 as part of the case study building which aims to provide additional in-depth insights into the experiences of digital collection developers. The participant for this case study has been selected based on the unique nature of collection and material that is being digitized and captured. The interview has been transcribed; the coding and analysis are ongoing.

We have been continuing 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 Gateway to Educational Materials (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 over 900 user keyword queries placed in the collection registry between February and September 2005 demonstrates a high level of subject searching made at the collection level (75% of the keyword searches). The most widely used user search categories were object, concept, place, and individual person, while the GEM subject scheme represents only broad concept category. We have also discovered the lack of semantic similarity between user queries and subject terms in controlled vocabularies, especially GEM. Only 2.2% of user search terms were matched in GEM subject scheme, while 26.25% were matched in Art and Architecture Thesaurus, and 72.6% -- 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.

Research Plans April 2007 through September 2007

In Spring and Summer 2007, we will interview Library Service and Technology Act (LSTA) grantees regarding this community's readiness to play in IMLS DCC 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). We would also like to obtain the data comparable to data collected from NLG grantees in terms of digital collection building challenges and successes. The sample has been selected; the documentation has been submitted to University of Illinois Institutional Review Board for review. We are planning to complete interview transcription, coding and data analysis by the end of September 2007.

We are also planning to continue comparative analysis of the perceived features of digital and physical collection based on the data obtained through interviews with digital collection developers conducted in 2004-2007.

¹ Please see Appendix 2 for figures summarizing major findings presented in this paper.

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 From Fair Use to Fair Trading: Creating a Digital Image Matchmaking Commons panel at the Visual Resources Association Conference, Issues of Collection Level Description and Its Relation to Item-Level Metadata in IMLS Digital Collection Registry and Item Repository, Discussion of RDA Panel at ALA Midwinter Conference, Dublin Core Metadata Initiative Standard (DCMI) revisions, Data Curation and Preservation, E-Prints Application Profile, NSDL Metadata Registry etc. 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>.

GEM

Work with the GEM subcontract continues according to the predetermined timeline. 33,567 records have been ingested into the GEM environment from the following collections: Illinois Alive! The Heritage and Texture of a Pivotal State During the First Century of Statehood (1818-1918); Historic Pittsburgh Image Collections; InfoMine; American Journeys; Louisiana Digital Library; Charles W. Cushman Photograph Collection; The Louisiana Purchase: A Heritage Explored; Columbia River Basin Ethnic History Archive.

Preliminary recommendations in regard to interoperability of IMLS DCC item-level metadata in an educational environment include:

1. The audience field should only be used if the resource explicitly states an audience type;
2. The audience field should not be used to drive traffic;
3. Collection-level records will be easier to adapt for use in The Gateway (as opposed to item-level records);
4. Collection and item-level metadata need to be significantly augmented to be useful in K-12 context;
5. Resources should be contextualized for use in a classroom.

Remaining IMLS DCC item-level records were supplied to GEM and will be ingested during the next performance period. Additionally, during the next performance period statistical analysis and long-term strategies to support educational use will be determined.

Certification

In submitting this report, I certify that all of the information is true and correct to the best of my knowledge.

Name and title of the person submitting the report and date of submission:

Timothy W. Cole, Principal Investigator

April 3, 2007

Amy Jackson, Project Coordinator

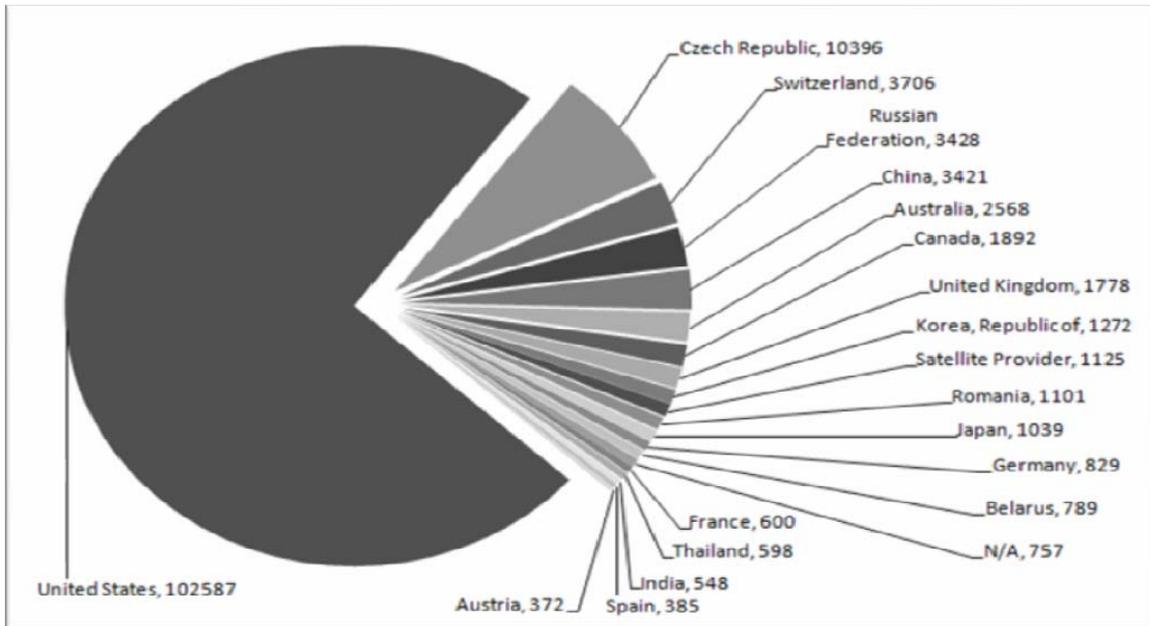
April 3, 2007

Appendix 1: IMLS Website Analysis

Overview:

From 09.01.2006 to 02.21.2007 the IMLS website received approximately 22,000 visitors from roughly 118 countries.

Top 20 Countries by Hit Count :



Search Engine Referrers:

Search Engine	Hits
Google.com	2715
Yahoo.com	85
MSN.com	50
Netscape.com	18
AOL.com	17
Ask.com	16
MyWay.com	9
earthlink.net	6
Dogpile.com	5
Websearch.com	4
Altavista.com	3
tiscali.co.uk	2
Vivisimo.com	1
scirus.com	1

Search Engine Keyword Referrers:

Keyword(s)	Hits
imls	63
ymca posters	24
imls dcc	19
imls digital collections registry	17
grant project	12
ymca poster	10
monoa public library	10
struwel peter	9
topography in london	9
topography of london	9
history of vocational education	8
shawnee indian beliefs	8

Appendix 2: Trends in Metadata Practices²

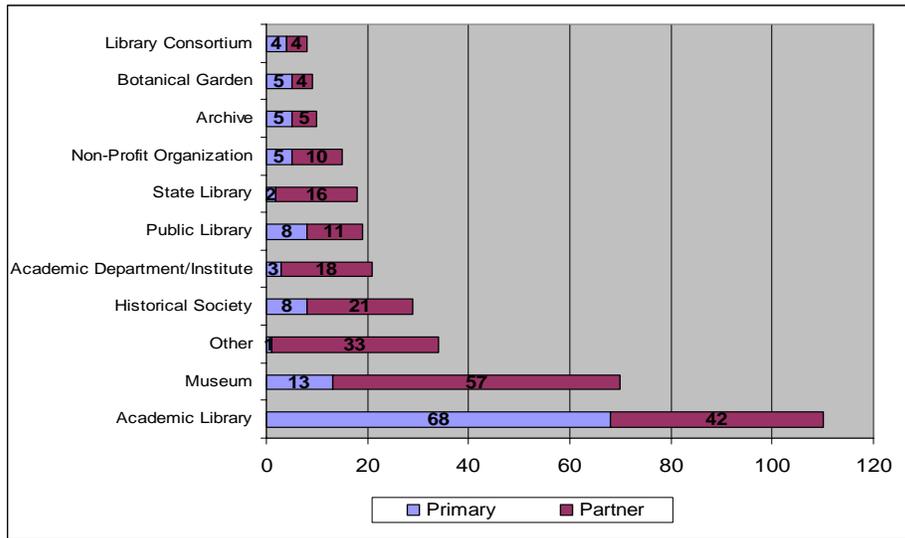


Figure 1. Number of Institutions Participating in the Registry by Type, 2003

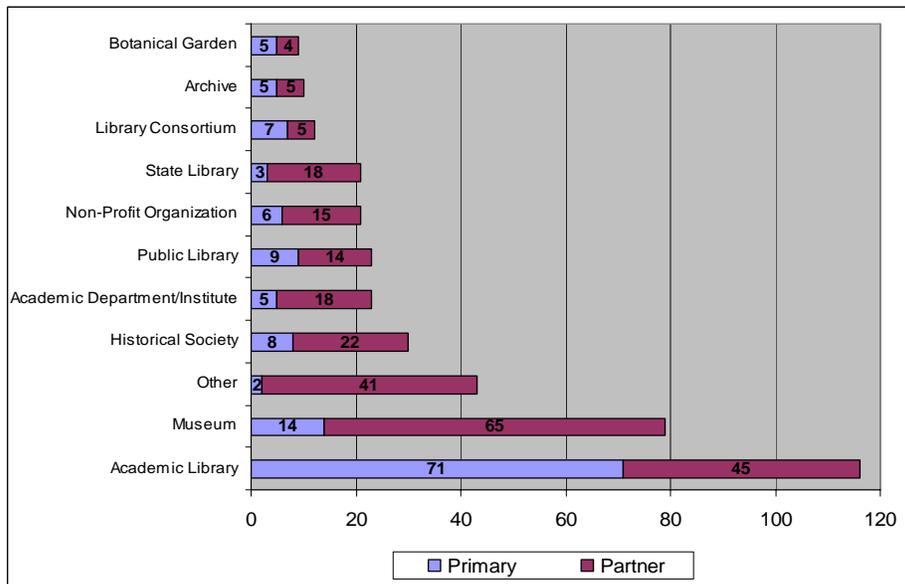


Figure 2. Number of Institutions Participating in the Registry by Type, 2006

² Figures are based on comparative analysis of 2003 and 2006 survey data presented in: Palmer, C., Zavalina, O., & Mustafoff, M. (2007). Trends in Metadata Practices: A Longitudinal Study of Collection Federation. Submitted to JCDL'07 Conference.

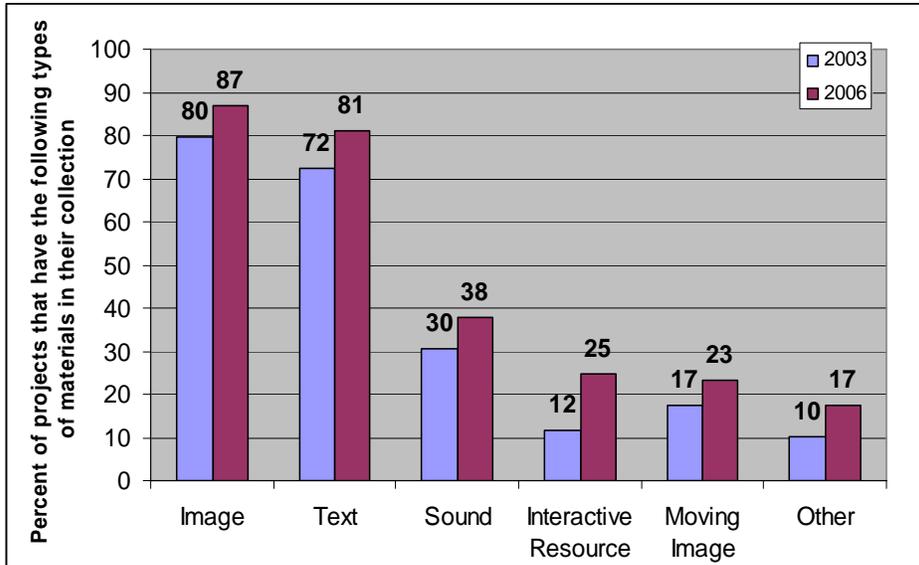


Figure 3. Types of Material in Collections, 2003-2006, (n=69)

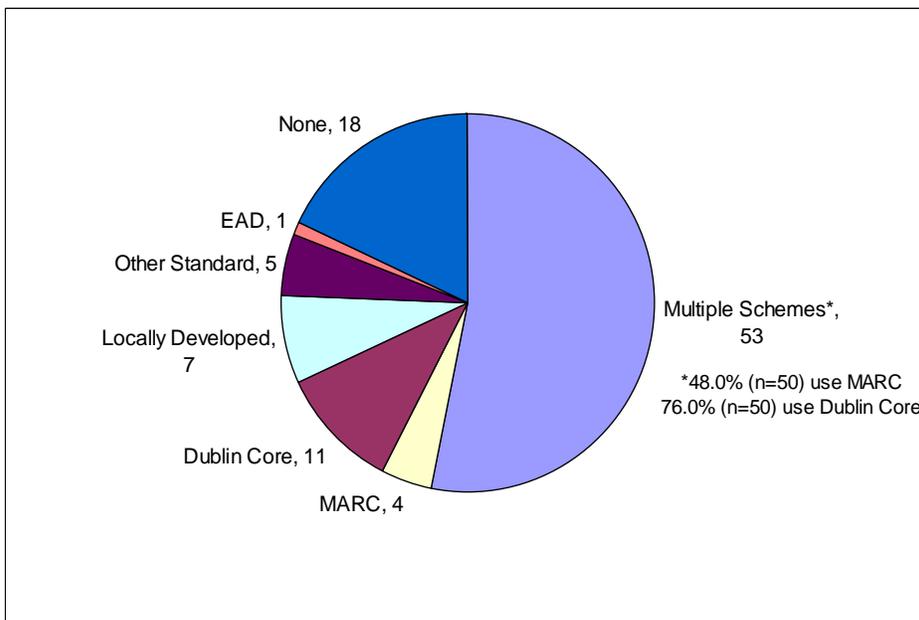


Figure 4. Percentage of Projects Using Single or Multiple Metadata Schemes for Item Level Description, 2003 (n=94)

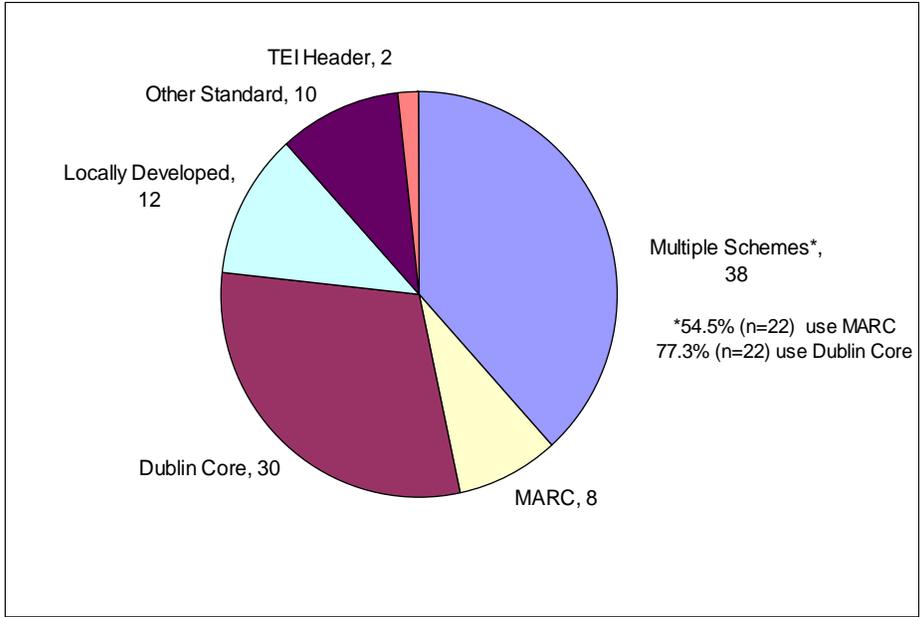


Figure 5. Percentage of Projects Using Single or Multiple Metadata Schemes for Item Level Description, 2006, (n=60)

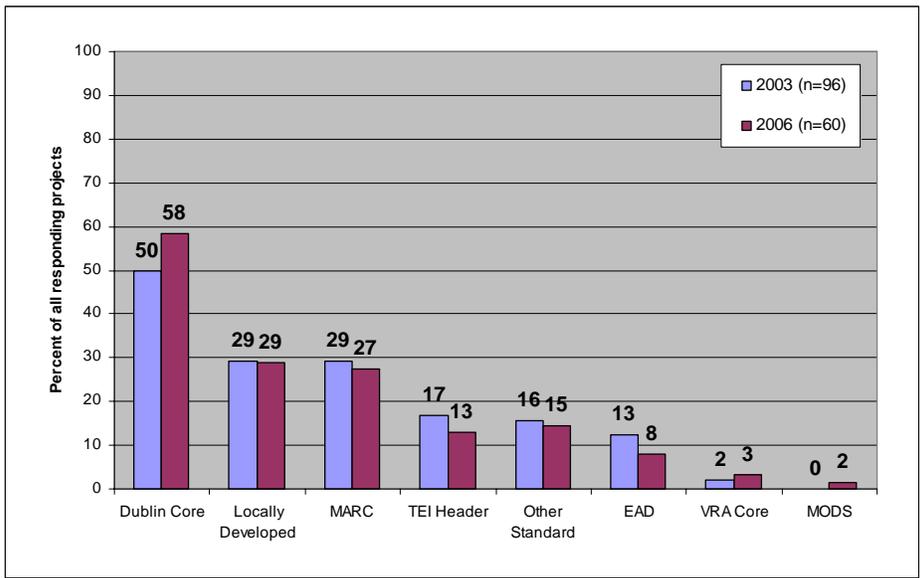


Figure 6. Percentage of Projects Using Metadata Schemes for Item Level Description (multiple schemes broken out), 2003-2006

Appendix 3: Steering Committee meeting notes

Steering Committee meeting notes

March 1, 2007

Hyatt Regency, Washington, DC

Present: Liz Bishoff, Tim Cole, Anne Craig, Martha Crawley, Diny Golder, Amy Jackson, Elizabeth Liddy, Carole Palmer, Paul Marty (by phone), Sam Quigley, Sarah Shreeves, Stuart Sutton, Katherine Wisser

The IMLS Digital Collections and Content Steering Committee met with the project team during WebWise 2007. The purpose of the meeting was to inform the Steering Committee of the current status of the IMLS DCC project, and to solicit feedback regarding potential future directions and broader impact of the projects.

Meeting notes:

Website:

- Discussed site design changes from 2006.
- Reported on usability studies and resulting interface changes (removal of search box from results page, collection results on left side instead of right side).
- Comments: Finding and exploring are the most demanding tasks—incorporate these tasks into future studies.

Website usage:

- Action item: OAI-PMH urls will be submitted to Google site maps.
- Action item: Check statistics for referrer pages—how many users are referred from IMLS?
- Recommendation: Ask that IMLS projects listed in the registry link to us.

Metadata trends:

- Comments: Results to date are inconclusive.
- Comments: Results may be obscured. Not all types of institutions will be using the same fields.
- Recommendation: Additional exploration of trends by institution type is warranted.

Recommendations from GEM

- Audience field should only be used if the resource explicitly states an audience type.
- Audience field should not be used to drive traffic.
- Collection-level records will be easier to adapt for use in The Gateway (as opposed to item-level records).
- Collection and item-level metadata need to be significantly augmented to be useful in K-12 context.

- Resources should be contextualized for use in a classroom.

Additional comments from Steering Committee

- How are users using the bookbag feature and creating personal collections? Can we enhance this feature?

Steering Committee Recommendations:

- Include metadata training with outcomes based training for both museums and libraries.
- Circulate list of questions to Steering Committee members for the final report and recommendations IMLS.

The IMLS Digital Collections and Content team would like to thank the Steering Committee members for their help and expertise through the course of the project. A list of questions regarding IMLS DCC will be circulated in mid-April, and feedback will be incorporated into the project's final report.