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

Interim Performance Report 3
1 October 2003 – 31 March 2004

Submitted by Timothy W. Cole, Principal Investigator, and
Sarah L. Shreeves, Project Coordinator
University of Illinois at Urbana Champaign
April 2003

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 3
1 October 2003 – 31 March 2004

Submitted by Timothy W. Cole, Principal Investigator, and Sarah Shreeves, Project Coordinator
April 2004

Summary

The IMLS Digital Collections and Content (DCC) Project has made progress on a number of fronts over the past six months. The response rate for the survey distributed in September to 1998-2002 National Leadership Grant (NLG) projects with digital content was 76% for Survey 1 and 72% for Survey 2. The analysis of the survey continues, and a second distribution of the survey via the World Wide Web is planned for early May to NLG projects funded in 2003 and to the non-respondents to the first survey distribution. IMLS approved the collection description metadata schema to be used for the initial release of the collection registry (which now contains 84 preliminary records). Development of the collection registry entry/edit forms is nearly complete and will be tested in May. We have continued to work with IMLS funded institutions to set up Open Archives Initiative (OAI) metadata provider services, although we hope to increase this work in the next six months. As of April 2004, the IMLS item-level metadata repository contained 77,631 metadata records from 20 NLG projects. The research team has continued to interview participants from selected projects, conducted a focus group at the 2004 Web-Wise conference, and continues to host a bi-weekly metadata roundtable attended by library school students and faculty as well as library faculty.

General Project Activities

Timeline
We are generally within the revised schedule of completion as submitted with the Interim Performance Report 2 (for April – Oct 2003).

Dissemination
The IMLS DCC project and its associated protocols and standards have been presented and discussed in several forums:

Tim Cole and Sarah Shreeves presented the OAI Protocol for Metadata Harvesting (PMH) at a pre-conference workshop for the Web-Wise Conference on March 3rd, 2004. See: http://imlsdcc.grainger.uiuc.edu/OAI_Tutorial_WebWise.ppt.

Tim Cole presented the IMLS DCC project at the Web-Wise Conference on March 4th, 2004. See http://imlsdcc.grainger.uiuc.edu/Cole_DemoIntros_WebWise.ppt.

Tim Cole and Sarah Shreeves attended the Collection Description Schema Forum (http://www.ukoln.ac.uk/events/cdfocus-schema-forum/intro.html) in London, England on February 12th, 2004. Sponsored by UKOLN (formerly the UK Office of Library Networking), the forum provided a useful opportunity to learn how other organizations were working with collection description and to discuss next steps for the RSLP Collection Description Schema and for the Dublin Core Collection Description Application Profile. Tim Cole facilitated a break-out session for this group. Tim Cole and Sarah Shreeves also attended the JISC Terminology Services Workshop (http://www.ukoln.ac.uk/events/jisc-terminology/) in London, England on February 13th, 2004.

The IMLS DCC project has several dissemination activities confirmed and in planning for the next six months to a year:

- In April 2004, Sarah Shreeves attended the NISO Metasearch Initiative meeting as a member of Task Group 2, charged with developing collection and service descriptions for use within a metasearch context.
- Sarah Shreeves will be presenting a seminar on OAI at the Technology for the Rest of Us seminar series at Ohio State in May 2004.
- Tim Cole will present at the "The Economics of Digitization: Toward Sustainability and Institutional Collaboration", an Innodata Isogen symposium in cooperation with the Newberry Library in May 2004.
- Sarah Shreeves has submitted a synopsis of the IMLS DCC Collection Description Metadata Schema to the CD Focus, a UKOLN-based newsletter for collection description.
- The September 2004 special issue of Library Hi-Tech devoted to IMLS National Leadership Grant projects will include an article on the IMLS DCC project.
- Sarah Shreeves will be speaking on metadata and the IMLS DCC project to the American Association of State and Local History (AASLH) annual conference in September 2004.
- Besiki Stvilia, a doctoral student on the project, has submitted a paper on metadata quality to the European Conference on Digital Libraries (ECDL) for September 2004. In addition, Tim Cole, Bill Mischo, and Tom Habing have submitted a paper on the OAI data provider registry.
- The American Society for Information Science and Technology (ASIST) has accepted a paper, "Metadata Practices and Implications for Federated Collections", from the research team for its annual conference in November 2004. In addition, Carole Palmer will also be presenting a paper on metadata standards adoption and knowledge-sharing on a panel at the ASIST conference.
- We will be presenting the experiences of the IMLS DCC project in a panel with Richard Rinehart of the Berkeley Art Museum and Martin Halbert of Emory University at the Museum Computer Network (MCN) annual conference in November 2004.

Steering Committee Activity
The Steering Committee met on March 5th, 2004 after the completion of the Web-Wise Conference in Chicago, IL to discuss survey results, research into metadata quality issues, and target audiences for the collection registry and item-level metadata repository. See http://imlsdcc.grainger.uiuc.edu/steeringcommittee/Mar5ActionItems.htm for action items from this meeting. (requires password).

Collection Registry Metadata Schema and Service

Survey of IMLS NLG Projects

Surveys 1 and 2 were sent in September 2003 to NLG projects. We created SQL databases (for each survey) to record the results of the surveys as they were returned. We began contacting non-respondents in early October and had achieved a 76% return rate on Survey 1 and a 72% return rate on Survey 2 by the end of March 2004. The research team is following up on the survey results with emails and phone calls.

Survey 1 has yielded some interesting results. 76% (50) of the respondents have divided their IMLS funded collection into sub-collections based on factors such as topic, administrative unit, type of material or a combination of these. 86% (57) of respondents did have item level metadata for the content in their digital collections. Of these 61% (35) were using multiple schemas. Most IMLS funded collections contain a combination of material types. Only 20% (13) of the 64 respondents to the material type question had a single material type in their digital collection. 34% (22) had a combination of image and text. Appendix One includes a synopsis of these and other Survey 1 results.

We have received Institutional Review Board approval for the web version of Survey 1 and 2 and are now in the process of setting up accounts to distribute the surveys to the 2003 NLG recipients (as indicated by IMLS) and to the non-respondents to the first wave of surveys. The responses to Survey 1 will contribute to the collection registry.

Collection Description Metadata Schema

The IMLS DCC Collection Description Metadata Schema was approved by IMLS in January 2004. We released it on our web site in February 2004. See Appendix Two for an overview of the schema. This is the working version of the schema, and minimal changes will be made only to bring it into alignment with international standards efforts. We expect to publish a crosswalk to other metadata schemas within the next six months. Other activities include developing an XML schema to express the metadata schema.

Sarah Shreeves has continued to be an active participant in the Dublin Core Collection Description Working Group, and has recently joined the NISO Metasearch Initiative Collection Description Task Group.

Development of Collection Registry

The results of the survey were used to populate the collection registry database developed to reflect the IMLS DCC Collection Description Metadata Schema. 84 collection records were created from the survey results and then edited and expanded through information gleaned from collection websites and other communications. Entry/edit forms have been developed and now being tested. A browse interface for the preliminary records was developed and is available at http://imlsdcc.grainger.uiuc.edu/collections/. We expect to submit the entry forms and schema for approval by the Office of Management and Budget in June. We have asked for feedback from IMLS about technology and accessibility requirements.

Item-Level Metadata Repository

Assisting projects in implementing OAI-data provider services

The survey results and continued discussions with NLG recipients about implementing OAI-data provider services have given us an clearer picture of the landscape. The breakdown of 1998-2002 NLG projects (95 total) in relation to OAI-data provider services is:

Category of 1998-2002 NLG Projects:	Number / % of NLG Projects:
Group 1 – Projects with OAI data provider sites for NLG content	21 (22 %)
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	21 (22 %)
Group 3 – Projects who meet certain technical criteria – e.g. have item-level metadata and a maintained web site	23 (24 %)
Group 4 – Projects with no item-level metadata, no interest in providing metadata via OAI, or whose grants were given up	13 (14 %)
Unknown	17 (18 %)
Total	95

We are currently working with Richard Rinehart at the Berkeley Art Museum/Pacific Film Archive (BAMPFA) to develop an ASP OAI-data provider service for his NLG project which utilizes FileMakerPro. Once installed this data provider will expose both the metadata from the IMLS-funded MOAC project (LL-90130).

In addition, the project created an OAI data provider for the IMLS-funded Illinois Alive project (LL-80052). The Illinois Alive collection consists of a series of web pages about Illinois history. Dublin Core metadata for each web page is embedded in the header of each. The IMLS DCC team developed a spider which crawled through the Illinois Alive pages to collect the Dublin Core metadata and store it within a SQL database on a server at UIUC. The metadata is then exposed via the OAI protocol. This implementation is a good example of a low-barrier entry into OAI for projects with metadata embedded within web pages.

We have continued to contact NLG recipients about implementing OAI-data provider services. As mentioned in Interim Performance Report 2 we continue to document some of the barriers to implementation. In particular we have found that if the technical infrastructure is in transition, institutions are reluctant to devote staff or time to implementation of OAI-data provider services. Our goal by the end of the project is to have approximately 50% of all NLG projects in our collection registry providing metadata via OAI. Over the next six months we plan to add fifteen to twenty projects to the metadata repository.

Metadata harvesting and design of item-level repository

We have continued to harvest and index item-level metadata from NLG projects. We are currently harvesting 77,631 records from 20 distinct NLG projects. The repository is currently available at: http://imlsdcc.grainger.uiuc.edu/search/. Sites and number of records harvested as of April 23rd, 2004 is available in Appendix Three. We have licensed the product, Spotfire, for assistance in our analysis of item-level metadata. Spotfire is used within the NSDL for metadata analysis.

The current interface for the item level metadata repository is an internally developed interface. The project team is interested in exploring different interfaces for the item level metadata repository including the Scout Portal Toolkit (http://scout.wisc.edu/Projects/SPT/) and the iVia Internet Portal (http://infomine.ucr.edu/iVia/), two open source portal products, as well as University of Michigan's DLXS product, XPat (http://www.dlxs.org/products/xpat.html).

Research

Data Collection
As of April 27, 2004, 26 interviews have been conducted with participants from 17 project sites. Thirteen of the interviews have been conducted in the last six months. Transcription and analysis of the interviews are well underway. Our research plan indicated that we would interview 15-20 project sites. The project team will interview people at least two other sites within the next two months.

For Survey 2 we have received a 72% response rate. Based on our analysis of the surveys we sent out two email follow-up questions that focused on sub-collections and the adequacy of the metadata scheme selected for search. The email follow-up questions were sent in mid-March and as of April 27, 2004 we have received a 60% response rate.

Ellen Knutson conducted a focus group of National Leadership Grantees in March 2004 at WebWise in Chicago. This focus group was attended by 11 people representing 9 institutions and 8 projects. The participants included representatives from museums, libraries, botanical gardens, and a zoo. The focus group format was used to gather more comprehensive data on how digital collections are being used by used at different institutions and the implications for repository development.

For the next year our data collection efforts will continue to focus on use and users of the individual NLG digital collections. We have begun to identify sites for our in depth case studies and expect to partner with some institutions to conduct user surveys. In February, Ellen Knutson traveled to Colorado for a site visit and to conduct interviews to further develop that partnership and deepen our understanding of their work.

Dissemination of Research Results
A report of the results of the project to date by Carole Palmer and Ellen Knutson has been accepted as a contributed paper for the American Society for Information Science and Technology annual meeting. The paper, Metadata Practices and Implications for Federated Collections, will be presented in Providence, RI in November 2004. The emerging themes presented in the paper are attached to this report in Appendix Four. At the same conference, Carole Palmer will also be presenting a paper on the panel being sponsored by the Digital Libraries SIG and the History and Foundations of Information Science SIG. The panel, coordinated by Marija Dalbello at Rutgers, is entitled Diffusion of Knowledge in the Field of Digital Library Development: How is the Field Shaped by Visionaries, Engineers, and Pragmatists? Palmer's paper is on metadata standards adoption and knowledge-sharing. Tefko Saracevic and Dalbello will be discussing how digital library research and practice inform each other.

Metadata Quality
Besiki Stvilia, a GSLIS doctoral student, was hired as a research assistant in January, 2004 to work on metadata quality issues. He has begun a series of analyzes on the item level metadata supplied through OAI. He has submitted a paper on metadata quality to the 2004 ECDL.

Related Activities
We have continued our bi-weekly metadata roundtable where members of the Graduate School and Library and Information Science and the University Library community meet to discuss issues that surround the use and creation of metadata. Some of the topics we have discussed include metadata quality and outcomes of the UKOLN and JISC Conferences. We also had discussion of key papers, including: "Digital Library Aggregation Services" by Martha Brogan and "Open Archives Initiative - Protocol For Metadata Harvesting Practices of Cultural Heritage Actors" by Muriel Foulonneau and David Dawson. Regular participants include faculty and doctoral students from GSLIS as well as university librarians and other GSLIS visiting scholars.

Appendix One – Selected Results from Survey One

Number of surveys sent: 92 (representing 95 total projects)
Number of non-active projects identified through survey or other communication: 5
Number of surveys sent to "active" projects: 87
Number of respondents to Survey One: 66 (76%)

_____

Sub-collections:

Number of respondents to question: "Is this collection divided into sub-collections (for example, type of material or subject area)?": 66 (100%)

Number of respondents with sub-collections: 50 (76%)

Basis of sub-collection organization:	Number (%) of respondents with sub-collections:
Administrative unit only	6 (12%)
Topic only	10 (20%)
Type of material only	8 (16%)
Other basis only	8 (16%)
Based on two factors:	
Administrative unit and Topic	2 (4%)
Administrative unit and Type of material	1 (2%)
Administrative unit and Other	4 (8%)
Topic and Type of material	5 (10%)
Topic and Other	2 (4%)
Based on three factors:	
Topic, Type of material, and Administrative unit	4 (8%)

Selected 'other basis' responses:

- Could also be sub–divided according to certain aspects of the collection, e.g. Ill. State Board of Education Learning Standards or teachers' lesson plans
- Keywords
- Grade level (age) appropriateness
- Time period
- Audience sub–collections; examples: educators, journalists, historical researchers, commercial, staff, museum partners
- Donating Individual or organization
- Taxonomic (biology) description at species level; character sets; image collection

Number of respondents to question: "How many sub-collections are within your overall collection?": 45 (90% of respondents)

| Number range of sub-collections	Number (%) of respondents to question
2-5 sub-collections	17 (38%)
6-10 sub-collections	10 (22%)
11-15 sub-collections	5 (11%)
16-20 sub-collections	3 (7%)
21-30 sub-collections	1 (2%)
31-40 sub-collections	2 (4%)
41 or more sub-collections	3 (7%)
Unknown or not relevant	4 (9%)

_____

Item level metadata:

Number of respondents with item level metadata: 57 (86 %)
Number of respondents without item level metadata: 9 (14 %)

Number of respondents using just one metadata schema: 22 (39% of respondents with item level metadata)
Number of respondents using multiple metadata schemas: 35 (61% of respondents with item level metadata)
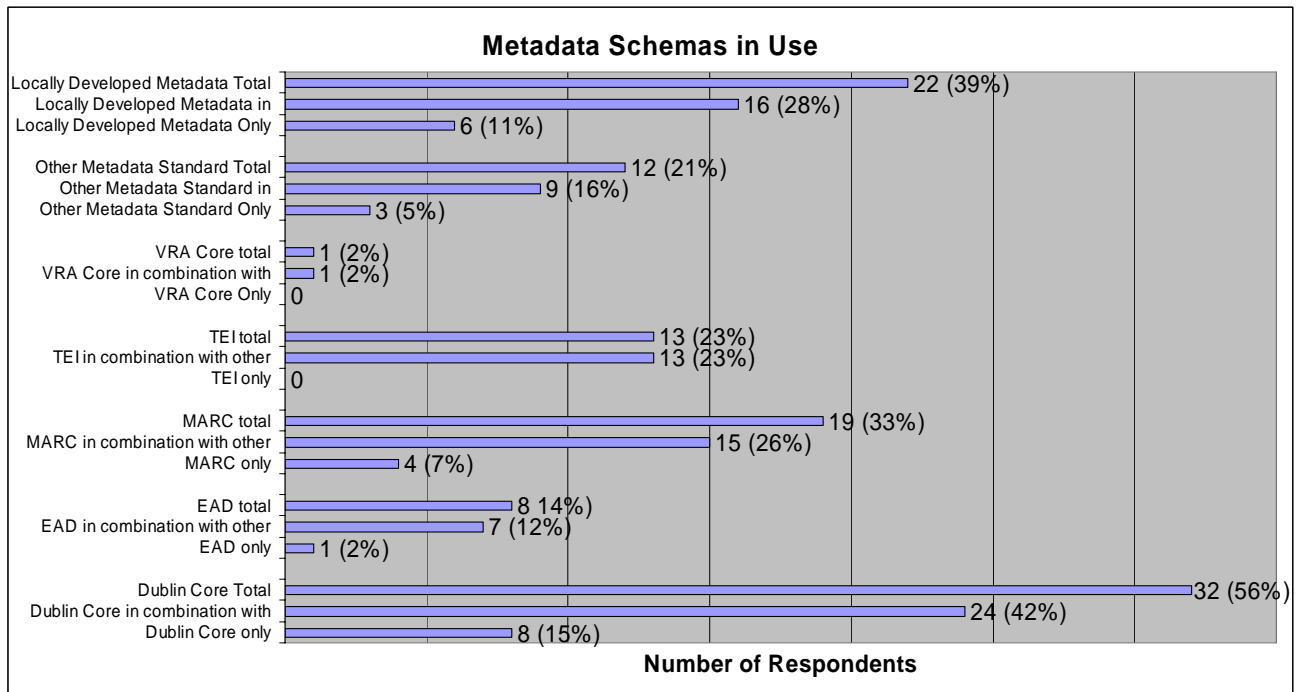
Other Standards in Use:
- Mets
- MOA2
- Museum MARC
- Darwin Core

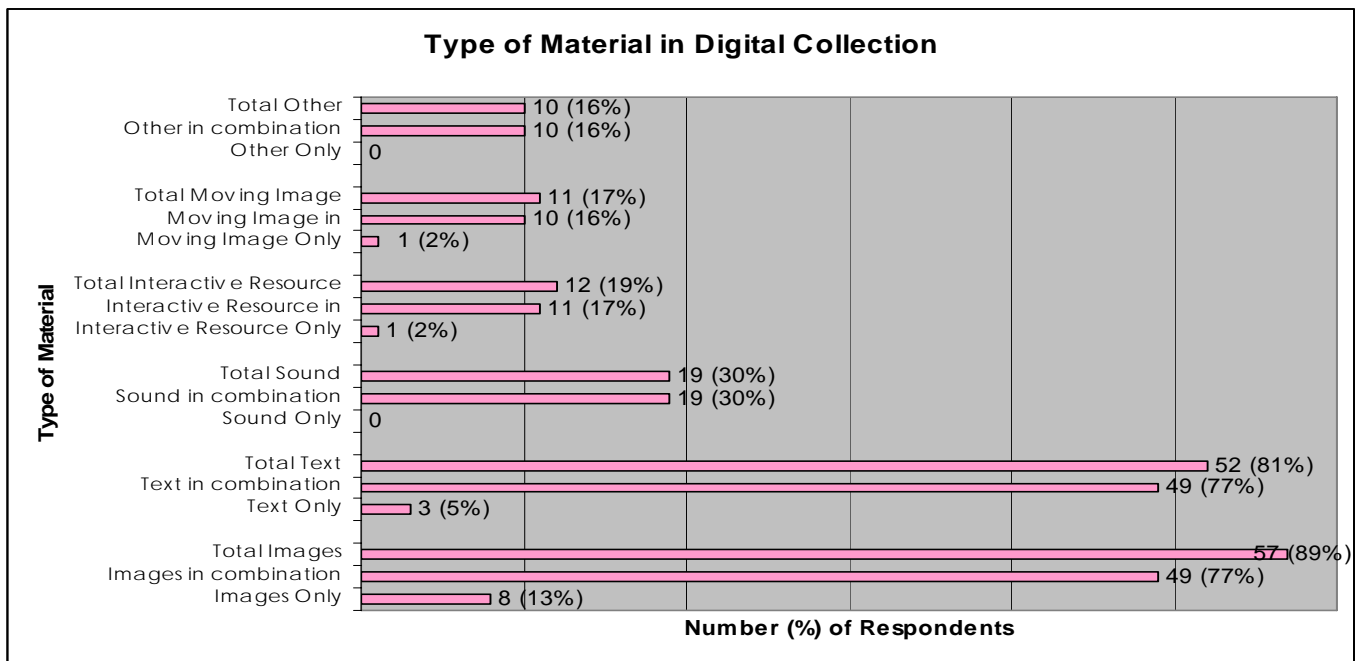- TDWG-SDD (Taxonomic Data Working Group - Structure for Descriptive Data)

_____

Controlled Vocabulary in Use:

Number of respondents who identified controlled vocabulary in use: 48 (84% of respondents with item level metadata)

Element	Top three used Controlled Vocabulary (% of respondents who identified C.V.)
Subject	LCSH (73%); LC TGM I (27%); AAT (17%)
Format	LC TGM II (17%); AAT (10%); MIME types (8%); AACR2 (8%)
Type	LC TGM II (21%); DCMI Type (13%); AACR2 (10%)
Personal names	LC Name Authority File (67%)
Geographic names	LCSH (27%); LC Name Authority File (25%); Getty Thesaurus of Geographic Names (15%)

_____

Type of Material In Digital Collection:

Number of respondents to question: "What type(s) of material have been digitized or created digitally?": 64 (97%)

Type of Material in Digital Collection

Type of Material (vertical axis) vs. Number (%) of Respondents (horizontal axis):

- Total Other: 10 (16%)
- Other in combination: 10 (16%)
- Other Only: 0
- Total Moving Image: 11 (17%)
- Moving Image in: 10 (16%)
- Moving Image Only: 1 (2%)
- Total Interactive Resource: 12 (19%)
- Interactive Resource in: 11 (17%)
- Interactive Resource Only: 1 (2%)
- Total Sound: 19 (30%)
- Sound in combination: 19 (30%)
- Sound Only: 0
- Total Text: 52 (81%)
- Text in combination: 49 (77%)
- Text Only: 3 (5%)
- Total Images: 57 (89%)
- Images in combination: 49 (77%)
- Images Only: 8 (13%)

Other Material Types in Digital Collections:
- Flash 'movies' (.swf)
- 3-D materials
- Artifact images, historic site views

- – 3-D Objects eg: plates, buttons, hats, mugs, jewellry, ribbons, lanterns, textiles, pens, trinkets, ceramics, etc.
- – botanic (herbarium) speciments
- – maps, music scores, 3-D artifacts (photos)
- – Atlas search
- – Currency
- – maps
- – illustrations within texts, fold out maps, charts, book jackets

Three most common combination of material types:

Combination of Material Type	Number (%) of respondents
Image and Text	22 (34%)
Image, Text, and Sound	5 (8%)
Image, Text, Sound, and Moving Image	4 (6%)

Other Observations from Survey:

Access restrictions:
98% (65) of respondents answered the question: "Is access to your collection limited to a specific group(s) of users?". Only one respondent placed access restrictions on their collection – and then only on a portion of their collection – specifically, copyrighted materials that could only be used for educational purposes.

Tracking use of collections:
95% (63) of respondents answered the question: "Is your project tracking usage of your digital collection through transaction log data?". Of these, 90% (57) were tracking use of their collections OR were planning to.

Collections developed prior to IMLS grant period:
97% (64) of respondents answered the question: "Was any digital content in collection developed prior to the NLG award?". Of these, 55% (35) had developed content prior to receiving the NLG award.

Continued development of collections after the IMLS grant period:
97% (64) of respondents answered the question: "Has/will digital content be added to collection after the completion of the grant period?". Of these, 78% (50) indicated that they would continue to add content to the collection, although 5 of these noted that additions would depend on additional resources (other grants, etc).

Appendix Two - IMLS DCC Collection Description Metadata Schema

The IMLS DCC Collection Description Metadata Schema is based on the UKOLN RSLP Collection Description Metadata Schema (http://www.ukoln.ac.uk/metadata/rslp/) and the Dublin Core Collection Description Application Profile (http://dublincore.org/groups/collections/). The IMLS DCC project has adapted these schemas to reflect the particular nature of the project and to incorporate the needs of National Leadership Grant projects. It is meant to describe digital collections created through IMLS funded National Leadership Grant projects and does not describe in detail the projects themselves. This metadata schema forms the basis of the IMLS NLG Collection Registry which is currently in development.

There are four entities described by the schema: the collection; the NLG project; the institution; and the administrator. A collection may have been created by multiple NLG projects and have multiple administrators. A collection may have only one hosting institution, but may have multiple contributing institutions. A collection may have multiple sub-collections, associated collections, or source physical collections. A NLG project may have only one administering institution, but may have multiple participating (or collaborating) institutions. The diagram below illustrates the relationships between these entities.
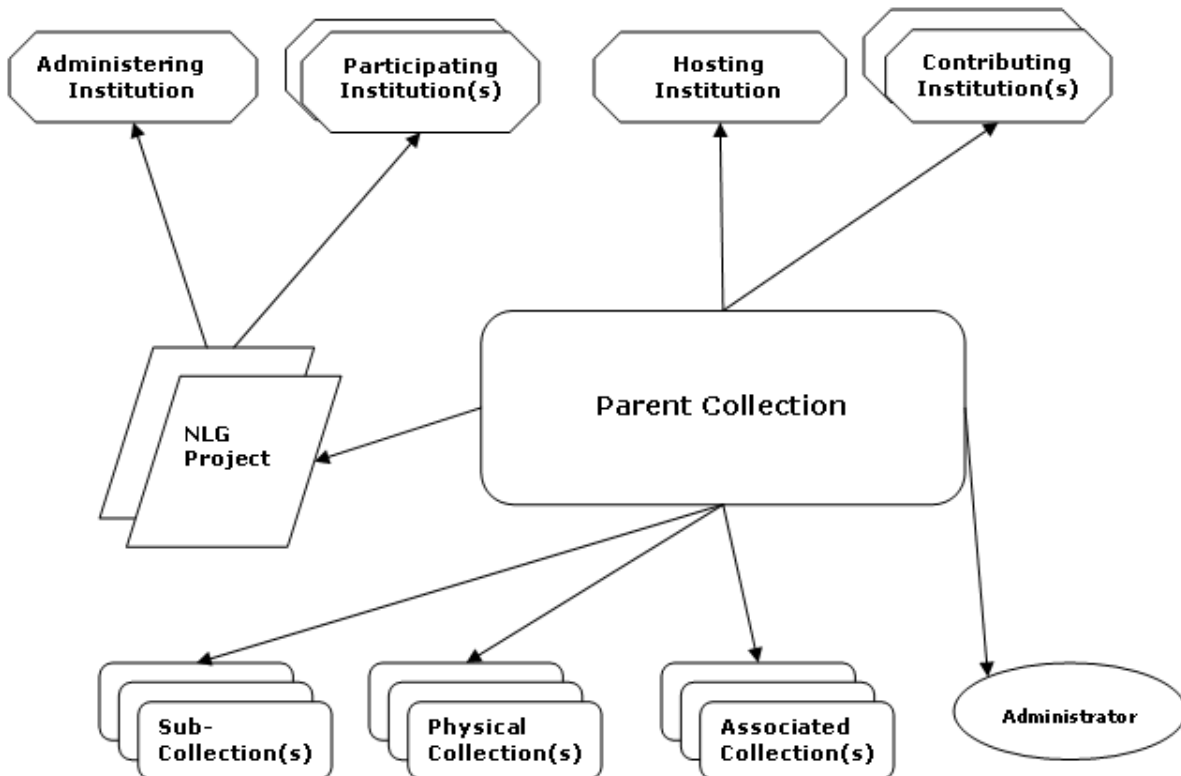
The IMLS DCC Collection Description Metadata Schema reuses metadata elements from other schemas and adapts their use. We have also developed elements that are needed for our purposes, but have mapped these to other standards as is possible (our crosswalk is coming soon). Included is a list of the vocabularies and namespaces used within the IMLS DCC CD schema.

Collection Entity

General attributes:

Display Name	Property Name	Definition
Title	dc:title	The name of the digital collection created by the National Leadership Grant project.
not displayed	dc:identifier	The unique key for the digital collection.
URL	imls:URL *sub-property of cld_gen:isAvailableAt *	The URL where an end user could access the digital collection / primary entry point for the digital collection.
Interaction with Collection	imls:interactivity	An indication of how a user can interact with the digital collection, such as search, browse, exhibit, etc.
Objects Represented	dc:type	The name of the object(s) or resource(s) represented by the digital item(s).
Format of Digital Items	dc:format	The format(s) of the digital items within the collection.
Size of Collection	dcterms:extent *sub-property of dc:format*	The number of digital items within the collection.
Frequency of Additions	imls:accrualFrequency *sub-property of cld:accrualStatus*	A statement of how often the digital collection is updated.
Language	dc:language	If textual, the language(s) of the items in the digital collection.
Audience	dcterms:audience	The primary audience(s) for the digital collection.
Supplementary Materials	imls:supplement *sub-property of dc:relation*	Materials such as lesson plans, docent materials, or exhibits that supplement the digital collection.
Access Restrictions	dcterms:accessRights *sub-property of dc:rights*	A statement of any access restrictions placed on the digital collection.
Rights	dc:rights	Information about rights (copyrights, etc.) held in and over the digital collection.
Collection Development Policy	imls:collectionPolicy *sub-property of dc:description	A statement of the collection development policy for the digital collection.

Alternative Access	cld_gen:isAvailableAt	The service that provides access to the digital collection, such as an OAI data provider or a Z39.50 target.
Metadata schema used	imls:metadataSchema	The name of the metadata standard(s) used to describe the items in the digital collection.
Notes	imls:notes	A statement of any additional information about the digital collection.

Topical attributes:

Display Name	Property Name	Definition
Topic	dc:subject	Terms that describe the overall topical content of the items in the digital collection. *Note: The IMLS Collection Registry will require the use of at least one GEM topic.
Description	dc:description	A summary of the content and topics of the digital collection.
Geographic Coverage	dcterms:spatial *sub-property of dc:coverage*	A place(s) or area(s) associated with most or all of the items in the digital collection.
Temporal Coverage	dcterms:temporal *sub-property of dc:coverage*	A time period(s) associated with most or all of the items in the digital collection.

Attributes describing relationships with other collections:

Display Name	Property Name	Definition
Parent Collection	dcterms:isPartOf *sub-property of dc:relation*	Any other collection(s) that contains the current collection.
Sub-Collection	dcterms:hasPart *sub-property of dc:relation*	Any other collection(s) contained within the current collection.
Source Physical Collection	dc:source	The physical collection(s) from which the current digital collection is derived.
Other Associated Collection	dc:relation	Any other collection(s) that is associated with or complements the current collection.

Attributes describing relationships with projects, institutions, and administrators:

Display Name	Property Name	Definition
NLG Project	dc:creator	The NLG project(s) which created the digital collection. [Link to the Project Entity]
Hosting Institution	dc:publisher	The institution which hosts (makes available) the digital collection. [Link to the Institution Entity]
Contributing Institution	dc:contributor	The institution(s) which have contributed content to the digital collection. [Link to the Institution Entity]
Administrator	imls:managedBy	The person who has primary responsibility for the digital collection. [Link to the Administrator Entity]

Project Entity

General attributes

Display Name	Property Name	Definition
Project Name	dc:title	The title of the IMLS National Leadership Grant project associated with the digital collection
IMLS Grant Number	dc:identifier	The IMLS grant number assigned to the NLG project.
Project URL	imls:URL *sub-property of cld_gen:isAvailableAt *	The URL where an end user can access the NLG project web site.

Attributes describing relationships with institutions:

Display Name	Property Name	Definition
Administering Institution	imls:administeredBy *sub-property of dc:relation*	The institution which has administrative responsibility for the NLG project (i.e. the applicant organization on the grant proposal). [Link to the Institution Entity]
Participating Institution	imls:participatedBy *sub-property of dc:relation*	The institution(s) participating in the NLG project either formally (with a collaborative agreement) or informally. [Link to the Institution Entity]

Institution Entity

Display Name	Property Name	Definition
Institution Name	dc:title	The name of the institution(s) associated with the digital collection and NLG project.
Not displayed	dc:identifier	The unique identifier for the institution.
Institution Type	dc:type	The type of institution, such as academic library, museum, etc.
State	imls:state	The state where the institution is located.
Zip Code	imls:zipCode	The zip code for the institution.

Administrator Entity

Display Name	Property Name	Definition
Administrator Name	vcard:fn	The name of the administrator responsible for the digital collection.
Not displayed	dc:identifier	The unique identifier for the administrator.
Email address	vcard:email	The electronic mail address for the administrator.
Affiliation	vcard:org	The institution the administrator is affiliated with.

Schemas (and Namespaces where available) used

cld_gen	General Description Terms identified in DC Collection Description App Profile	reference: http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/2004-02-01/
cld	Collection Description Terms identified in DC Collection Description App Profile	reference: http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/2004-02-01/
dc:	The Dublin Core Metadata Element Set v. 1.1	namespace: http://purl.org/dc/elements/1.1/ dc
dcterms:	Dublin Core Terms	namespace: http://purl.org/dc/terms/dcterms
imls:	IMLS DCC Collection Description Metadata Schema	reference: http://imlsdcc.grainger.uiuc.edu/CDschema_elements.htm
vcard:	RFC 2426 - vCard MIME Directory Profile	reference: http://www.ietf.org/rfc/rfc2426.txt

Appendix Three – National Leadership Grant Collections and Number of Records Harvested

77,631 metadata records from 20 distinct NLG projects

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
111 records

Colorado Digitization Program
"Heritage Colorado" - LL-90094
27,297 records

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

Louisiana State University
"Louisiana Purchase Bicentennial: A Heritage Explored" - ND-00010
539 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
3342 records

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

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

University of Illinois

"Teaching with Digital Content" - NL-00003
1,544 records

University of Maine
"Maine Music Box" - LG-03-02-0116
7,420 records

University of Michigan
"Flora and Fauna of the Great Lakes" - NL-00034
32,766 records

University of Minnesota
"Summons to Comradeship: World War I and II Posters" - ND-10007
2306 records

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

University of North Carolina
"The North Carolina Experience: Beginnings to 1940" - ND-00031
431 records

University of Tennessee
"Tennessee Documentary History" - ND-10020
1,207 records

University of Tennessee
"Frank H. McClung Museum WPA/TVA Photograph Archive" - LG-03-02-0080
1480 records

University of Wisconsin-Madison
"Africa Focus" - LL-80131
3650 records

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

Appendix Four – Emerging Themes from Surveys, Proposals and Interviews

Ellen Knutson, Carole Palmer, Michael Twidale

BASELINE:

Types of Institutions: Although there are a variety of institutions that have received funds from IMLS for digitization projects, by far the most common type is academic libraries. In fact, only 29 out of 122 projects did not involve an academic library, academic department, or a museum based in a university.

Types of collections: It perhaps comes as no surprise that the vast majority projects are digitizing images of artifacts, maps, photographs, museum objects, and different kinds of texts. While the image format limits search and analysis capabilities of text, it does allow users to see the documents in their original form and condition. A small percentage of the collections are exclusively texts, and a few projects are producing encoded texts, with about 20 using or investigating the use of TEI in their projects. It seems that there may be inconsistency in format and type description with text collections, so we will monitor this as the repository develops.

Metadata: MARC and Dublin Core are the schemes that are used most frequently either alone or in combination with other schemes. Less than one third of the all the projects did not use MARC or Dublin Core. 35% of the projects used or proposed to use multiple schemes. The use of multiple schemes did not correspond with collaborative projects.

METADATA:

Scheme selection: The degree to which a standard had been previously implemented and tested was of central importance, as was use by peer institutions. Compatibility with local systems was also a driving force. Several librarians reported that their choice of MARC was due to their OPAC's inability to handle Dublin Core records. Projects chose to apply a local scheme (or in some cases forego descriptive metadata altogether) for a number of reasons: customization was needed to capture information unique to the materials, information already recorded in a database or some other local information source was to be imported, or existing standards did not allow projects to adhere to their goals. In our interviews, participants expressed a preference for MARC's field richness, but valued Dublin Core for its perceived ease of application.

Field use: Nonstandard use of fields seemed to be more prevalent with Dublin Core. For example, in one case the source field was appropriated to provide information about the original object that had been digitized, and in other projects the data placed in the description field had been extended to compensate for the lack of appropriate fields in Dublin Core. Inconsistency in the date field was an issue that came up in several interviews.

Issues with applying scheme: The three most common problems with description were: consistent application of the chosen metadata scheme within a project, identification and application of controlled vocabularies, and integration of sets of data, schemes, and vocabularies either within an institution or among collaborators. Other issues that arose in applying the scheme that did not have to do with description per se include: standards development, finding qualified staff (though this was not a uniform response), technical problems especially with software.

Collaborative projects: The importance and utility of collaborative initiatives was strongly represented in the interviews, however collaboration did pose additional metadata challenges. Choosing a metadata scheme that works well across varied cultural heritage institutions is the first challenge, and there was no evidence of any scheme meeting the expectations or needs of all the institutions. Consistent application of the selected scheme was another ongoing difficulty, in part because of the distinct cultures of description that have evolved in different kinds of institutions.

USERS:

Audience: Most projects thought of their audience as a range of users. Some even going so far as to say: "The whole world."

Access: From our interviews it seems that resource developers feel that the digital representation provides better access to the material for their users.

Users will be one of our main focuses for the research in the up coming year, beginning with the focus group that was organized at WebWise.

REPOSITORY APPLICATIONS:

Collection description: The survey responses to a question about elements that should be designated for collection description proved to be well aligned with the DCC schema under development at the time. Suggestions for elements, beyond the basics of title, subject, description, type, format, etc., often reflected traditional modes for identifying collections in museums and archives, such as by donor or correspondent. Fields for user oriented data were also suggested, such as audience and lesson plans. Participants emphasized institution as a primary element for the collection description scheme, and some suggested all contributing institutions should be identified (which for some projects could be in the hundreds). Many also mentioned the desire to connect or situate the digital collection in the context of the physical collection.

Collection Definition: This has proved to be more contentious than we initially anticipated. Moreover, the content of some projects may not be considered collections, per se, by their developers, but rather exhibits, learning modules, or multimedia compilations. In the interviews, respondents frequently did not have a firm idea of how many collections they were creating, suggesting that they may not have yet thought about how their collections should be represented in a federated repository. A few avoided answering the question altogether. This excerpt is a good example of what we encountered: "We have a problem with that word collection. We fought about that word, so when you use it what do you mean?"

Anticipated Repository Use: Respondents' ideas about the NLG federated collection were amorphous. In both the surveys and interviews, many respondents were unsure of the role of the DCC repository. As might be expected from this group of respondents, there was considerable interest in the resource for information on up-to-date practices for digital projects and IMLS funding trends. But, clearly this type of current awareness could be achieved with a project directory and would not require building a repository. Forty percent of the respondents recognized how the resource could benefit reference and research services at their institutions, but few perceived it as a helpful tool for end users. There were scarcely any comments about the repository's potential for supporting programmatic resource sharing or the creation of new configurations of collections.