

Proposal for an IMLS Collection Registry and Metadata Repository

Timothy W. Cole, William H. Mischo, Carole L. Palmer, Michael Twidale, Nuala A. Bennett
University of Illinois at Urbana-Champaign

ABSTRACT

The visibility of a digital library collection and the ease with which individual items within such a resource may be discovered are increasingly important predictors of how widely and frequently collection content will be used. Although there are differences in the specific manner in which museums, libraries, and archives define and implement collection constructs, all traditionally make extensive use of such constructs to organize and delineate their holdings. In the digital world, where the risk of quantity overwhelming quality is high, the organization of content into collections is especially helpful. Properly designed collection registries can help to organize large aggregations of digital content from multiple institutions and make relevant resources easier to find and more visible to end-users.¹ Inclusion of collection-level metadata in a broadly comprehensive, searchable registry is a way to enhance the visibility of an online information resource. Sharing item-level metadata within a collection or repository has the potential to enhance the discoverability of individual items. Digital library developers are now promulgating architectural models and digital library service implementations that assume and require distributed, reusable repositories of content described by appropriate item-level metadata.² Long-term value and utility of digitized content is greatly enhanced through participation in collection-level registry services and, when appropriate to the nature of a collection, the implementation of item-level metadata sharing protocols.

The University of Illinois at Urbana-Champaign proposes to design, implement, and research a collection-level registry and item-level metadata repository service that will aggregate information about digital collections and items of digital content created using funds from Institute of Museum and Library Services (IMLS) National Leadership Grants. This work will be a collaboration by the University Library and the Graduate School of Library and Information Science. All extant digital collections initiated or augmented under IMLS aegis from 1998 through September 30, 2005 will be included in the proposed collection registry. Item-level metadata will be harvested from collections making such content available using the Open Archives Initiative Protocol for Metadata Harvesting (OAI PMH).³ As part of this work, project personnel, in cooperation with IMLS staff and grantees, will define and document appropriate metadata schemas, help create and maintain collection-level metadata records, assist in implementing OAI compliant metadata provider services for dissemination of item-level metadata records, and research potential benefits and issues associated with these activities. The immediate outcomes of this work will be the practical demonstration of technologies that have the potential to enhance the visibility of IMLS funded online exhibits and digital library collections and improve discoverability of items contained in these resources. Experience gained and research conducted during this project will make clearer both the costs and the potential benefits associated with such services. Metadata provider and harvesting service implementations will be appropriately instrumented (e.g., customized anonymous transaction logs, online questionnaires for targeted user groups, performance monitors). At the conclusion of this project we will submit a final report that discusses tasks performed and lessons learned, presents business plans for sustaining registry and repository services, enumerates and summarizes potential benefits of these services, and makes recommendations regarding future implementations of these and related intermediary and end user interoperability services by IMLS projects.

¹ Paul Miller. 2000. "Collected Wisdom: Some Cross-Domain Issues of Collection Level Description," *D-Lib Magazine*, vol. 6 (9). Available: <http://www.dlib.org/dlib/september00/miller/09miller.html>

² See, for example, Diane Hillmann's "Metadata Primer" for National Science Digital Library participants. Available: <http://metamanagement.com.nslib.org/outline.html>

³ "The Open Archives Initiative Protocol for Metadata Harvesting."

Available: http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html

Proposal for an IMLS Collection Registry and Metadata Repository

Timothy W. Cole, William H. Mischo, Carole L. Palmer, Michael Twidale, Nuala A. Bennett
University of Illinois at Urbana-Champaign

NARRATIVE

National Impact

That which makes for a successful library or museum digitization project has changed over the course of the last decade. Ten years ago proof of concept and the innovative application of technology to digitize primary content were sufficient objectives. Digitization projects today must also address concerns of reusability and interoperability. Added value is given to digital collections that can function as components and building blocks upon which advanced digital library services may be built. Our proposed project will provide additional infrastructure and experience for IMLS and its digitization project grantees. Inherently this project has the potential to enhance the value of prior IMLS projects and contribute to the success of future ones by introducing infrastructure designed to enhance reusability, interoperability, and discoverability. Interest in interoperability technologies continues to grow both nationally and internationally as government and private sector concerns alike seek ways to share resources while still maintaining ownership and authoring rights. Applying OAI PMH tools and structures to this IMLS project provides an opportunity to propose a model for other national scale projects interested in building similar repository services.

The ultimate impact of this project will come not just from the creation of an IMLS collection registry and metadata repository, but also from establishing the viability and usefulness of sharing collection-level and item-level metadata across a wide range of IMLS digitization projects. Recent research indicates both collection-level and the item-level metadata are essential for a full complement of online information services. In regard to the UK's Research Support Libraries Programme Collection Description Project, researchers report, "a strong view is emerging that libraries need to complement item-based description with description at a higher level." They go on to suggest that describing collections in a standard, well-structured fashion would better enable users to discover and locate resources and search across multiple collections. Additionally such well-structured collection descriptions would support the refinement of distributed searching algorithms and facilitate the creation of software to perform more precise distributed searches on behalf of users.⁴ Approaching the same issue from a museum-oriented perspective, Heather Dunn of the Canadian Heritage Information Network suggests, "collection-level descriptions facilitate cross-disciplinary, multi-level access to Web and database resources for a diverse audience."⁵ In regard to item-level metadata, Carl Lagoze and Herbert Van de Sompel, in introducing the OAI PMH in a paper presented at the 2001 Joint Conference on Digital Libraries,⁶ point to the benefits of item-level metadata sharing as demonstrated in Cornell's Networked Computer Science Technical Research Library (NCSTRL) project and similar projects elsewhere. NCSTRL, (e.g., Lagoze and Fielding, 1998)⁷ was an attempt to translate into distributed information space traditional library collection development functions in the form of a digital library "collection service." Within federated digital library systems featuring full-text searching, such as the testbed developed for the Illinois Digital Library Initiative / D-Lib Test Suite

⁴ Andy Powell, Michael Heaney, and Lorcan Dempsey. 2000. "RSLP Collection Description," *D-Lib Magazine*, vol. 6 (9). Available: <http://www.dlib.org/dlib/september00/powell/09powell.html>

⁵ Heather Dunn. 2000. "Collection Level Description - the Museum Perspective," *D-Lib Magazine*, vol. 6 (9). Available: <http://www.dlib.org/dlib/september00/dunn/09dunn.html>

⁶ Carl Lagoze and Herbert Van de Sompel. 2001. "The Open Archives Initiative: Building a low-barrier interoperability framework," in *Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries*, ACM Press: New York, 54-62. Available: <http://www.openarchives.org/documents/oai.pdf>

⁷ Carl Lagoze and David Fielding. 1998. "Defining Collections in Distributed Digital Libraries," *D-Lib Magazine* (November). Available: <http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>

project (1994 - 2001), item-level metadata has facilitated search normalization and result display standardization. It has also played a crucial role in enabling reference linking systems such as CrossRef's implementation of Digital Object Identifiers and the SFX implementation of OpenURLs.

A question that remains to be answered is what are the long-term benefits of collection registry and item-level metadata repository services for digital projects such as those sponsored by IMLS? Can we envision the ultimate end-users? Will it be K-12 teachers developing their next lesson plans? Will it be historians attempting to weave together threads from our collective pasts? The goals of this project will be accomplished by working with IMLS grantees to select and develop standards and best practices for disseminating metadata that describes IMLS funded collections. This work must be achieved in a manner that promotes maximum reusability and interoperability and proactively and aggressively encourages grantees to participate in metadata sharing. Selection of standards and the extension and customization of them for application within the diverse domain encompassing IMLS digitization projects is a significant challenge. The key to success will be to select and develop metadata standards, requirements, and guidelines that are consistent with the range of technical and descriptive capabilities and resources available to IMLS grantees and to develop a working plan that allows for the continual evolution of these standards. Our experiences during recent and current IMLS grant projects and during our current OAI metadata harvesting project suggest that innovative approaches will be called for to help IMLS grantees participate in metadata sharing, particularly at the item level. While collection management databases are becoming commonplace in most IMLS grantee institutions, relatively few are set up to support generalized cross-database metadata sharing. Many sites still have minimal, limited Web public access components in place. Network connectivity is still an issue for some. The metadata that is available often is more administrative and structural metadata than descriptive metadata suitable for end-user resource search and discovery.

Our approach will begin with the recognition that initial consultation and advice to potential metadata providers must be tailored on a case-by-case basis. Museums, libraries, and archives each have their own formats and standards for describing the range objects held by the institution. Not everyone speaks a common language, even in cases where two museums contain objects and information on the same or closely related subjects. Controlled vocabularies vary necessarily in their depth of coverage in order to accommodate various levels of description. In past and current projects we have worked with a wide range of museums, libraries, and archives examining their approaches to collection management and use of descriptive metadata content. We have examined the kinds and sources of controlled vocabularies used, and the completeness and precision of terms and conditions of collection access and use. Such case-by-case examinations of existing metadata provider information infrastructure is essential. Once the intellectual effort to clarify this infrastructure has been detailed and negotiated, the transformation, augmentation, and integration of available metadata into normal workflows can be designed for a particular repository built specifically with sharing functions in mind. While the resources available for this proposal would preclude on site re-engineering of each IMLS grantee's overall information system design, we can provide extensive guidance and expertise. On earlier projects this has included approaches such as capturing data in a repository's existing data management format and transforming it on our local servers at Illinois into a format compatible with OAI PMH. Scripts and transforming stylesheets developed during this process were provided to the data provider to facilitate implementation on the their site. In the interim a surrogate metadata provider were maintained on Illinois servers for initial testing and demonstration purposes.

Adaptability

Key to adaptability of technology developed in the course of this project will be our use of standard, ubiquitous technologies and the open publication of schemas, stylesheets, and developed software. OAI PMH is inherently based on ubiquitous technologies such as HTTP, XML, and the Dublin Core metadata schema. The University of Illinois has recently registered an open source software license with the Open

Source Initiative.⁸ For our current OAI metadata harvesting project, all software developed, including metadata harvesting and provider tools and transforming XML stylesheets, is available in accordance with this license. Additionally, an Illinois OAI project site has been created on the SourceForge.Net server.⁹ We propose to continue this site for the IMLS collection registry and metadata repository project. All software developed in the course of the project, including scripts and XSL-T stylesheets will be posted on this site and made available free of charge. Additionally, we propose to implement services and applications on ubiquitous platforms, including at least (but not limited to) Microsoft Windows 2000 / XP and Linux. While some commercial infrastructure server software would necessarily be used, selection of this software will be made consistent with a desire to insure widespread adaptability. With the current Illinois OAI project, development is done in parallel for both the Microsoft Internet Information Server and the Apache/Tomcat Web server. Database management software used is MySQL (freely available), XPat (University of Michigan Digital Library Extension Service), and Microsoft SQL 2000 and Access 2000. Software is written in Java, Perl, and/or VBScript (ASP). We propose to continue this development approach for the IMLS collection registry and metadata repository.

Design

The immediate beneficiaries of collection-level and item-level metadata sharing will be the participating IMLS grantees who realize the benefit of enhanced visibility and discoverability for their content. However, the magnitude of that benefit depends on the degree to which intermediate and ultimate end-users can profit from the aggregation of metadata. We recognize the importance of working closely with both these audiences. To engage the IMLS grantees, the collection description requirements and item-level metadata provider services implemented must be perceived to meet their needs, while at the same time being achievable within their means. Concurrently, the benefits of aggregating this information for end-users must be demonstrated to justify the cost. Our project proposal is designed to efficiently and effectively implement a collection registry and metadata repository service while investigating the nature and scope of benefits to intermediate and end-users. The work will build on an extensive body of prior work in metadata and digital library interoperability at Illinois and elsewhere.

Context. Illinois has been actively engaged with the OAI community since becoming an alpha tester of the PMH in the fall of 2000. Work at Illinois and elsewhere has created a body of tested open source metadata provider and harvesting software tools written for a variety of platforms and in a variety of programming languages. The focus of our Andrew W. Mellon Foundation funded OAI metadata harvesting project¹⁰ is on cultural heritage content, and this has given us an opportunity to investigate and develop OAI PMH resources optimized for the kind of metadata content typical of many IMLS digitization projects. We currently are working with metadata from more than 30 institutions and consortia, including museums such as the American Museum of Natural History and the University of Illinois Spurlock Museum, historical societies and archives such as the Minnesota Historical Society, the Ohio Historical Society, and the Bentley Historical Library, academic libraries such as Harvard University Library and the libraries of nine Committee on Institutional Cooperation (CIC) member institutions, governmental libraries and systems such as the Library of Congress and the Illinois State Library, and statewide digitization projects such as the Colorado Digitization Project and the Online Archives of California.¹¹ We have worked with metadata describing images, manuscript archives, historical artifacts, recorded history audio files, historical maps, and digitized texts. This experience

⁸ Available: <http://www.opensource.org/licenses/UoI-NCSA.html>

⁹ Available: <http://sourceforge.net/projects/uilib-oai>

¹⁰ Timothy W. Cole, et al. 2002. "Now That We've Found the 'Hidden Web,' What Can We Do With It? The Illinois Open Archives Initiative Metadata Harvesting Experience," in *Museums and the Web 2002: Proceedings*, David Bearman and Jennifer Trant (eds.), Archives and Museum Informatics: Pittsburgh: 63-72. Available: <http://www.archimuse.com/mw2002/papers/cole/cole.html>

¹¹ Complete list of metadata providers available: <http://oai.granger.uiuc.edu/AboutCollections.htm>

yields insights into the information environments of a disparate group of cultural memory institutions and the issues that arise when trying to aggregate metadata provided by them. Our proposal builds on this body of experience.

Illinois IMLS funded projects such as Digital Cultural Heritage Community (DCHC) and Teaching with Digital Content (TDC)¹² give us experience with a range of museums and libraries actively involved in digitizing content and with user communities such K-12 teachers. This work provides insights into the needs and expectations of both metadata providers and potential users. An important observation from this work is the recognition that museums tend to be more focused than libraries on the interpretation of materials. This suggests collection-level and item-level metadata schemas need to include information that will aid in interpreting the significance and relationships of collections and artifacts described. For instance, the metadata schemas developed for our DCHC and TDC projects added fields to contain this information. The added interpretation fields assisted teachers in making the information and artifacts described come alive in the classroom. (Our extension of the version 1.1 Dublin Core metadata schema anticipated the recent decision by the Dublin Core Metadata Initiative to add an "audience" element to the Dublin Core schema.) Museum curators and teachers alike agreed that the repository provided by our projects offered an opportunity to explore new models of partnerships at regional, state and national levels. Teachers suggested the inclusion of an interface component for submitting commentary on objects. Curators and librarians welcomed this suggestion and indicated they too would find it useful to receive feedback from users about objects, documents, interpretations, and descriptions. Teachers commented that in the past, it was difficult to find historical and social science primary source material because there was so much information available. They liked the projects' metadata repositories because they put usable information in a central, more trusted location.

Other prior work involving digital information resources¹³ provides a source of expertise and familiarity with large scale SGML and XML applications, metadata generation and manipulation, metadata schema development and documentation, and end-user search and retrieval interface design. This work included participation in the DOI-X project in the fall of 1999, which led to the establishment of the CrossRef consortium in early 2000. More recent work on another project¹⁴ focused the use of database resource-level descriptions to help end-users navigate between and among online bibliographic databases.

Implementation of Collection Registry and Metadata Repository Services. Based on this range of experience, knowledge, and expertise, we propose to implement the collection registry and metadata repository service by performing the following tasks.

- Survey IMLS grantees to ascertain the nature and scope of their collections and availability of item-level metadata and supporting information such as project transaction logs, and user studies.
- Define a collection-level metadata schema. Perform literature review and search for available collection-level metadata schemas. Select and refine as necessary (in concert with IMLS) a schema appropriate for use in constructing a collection registry.
- Implement collection registry service. Work with IMLS grantees to create and maintain in the registry descriptive collection-level metadata records for each extant collection created as part of an IMLS

¹² Nuala Bennet, et al. 2002. "Illinois Digital Cultural Heritage Community – Collaborative Interactions Among Libraries, Museums, and Elementary Schools." *D-Lib Magazine*, vol. 8 (1). Available:<http://www.dlib.org/dlib/january02/bennett/01bennett.html>

¹³ Timothy W. Cole, et al. 2001. "Using XML and XSLT to Process and Render Online Journals," *Library Hi Tech* 19 (3): 210 - 222

¹⁴ Wei Ma. 2002. "A Database Selection Expert System Based on Reference Librarian's Database Selection Strategy: A Usability and Empirical Study," *Journal of the American Society for Information Science and Technology* vol 53 (7): 567-580.

National Leadership Grant projects initiated 1998 to present, adding new collections as they go online through September 30, 2005.

- Ongoing testing and refinement of collection registry service based on outcome of research investigations and feedback solicited from IMLS, participating projects, and selected end-user groups.
- Analyze available item-level metadata and metadata schemas. Identify any supplemental community specific schemas for use with this project. Identify transformation services and OAI metadata provider tools that are necessary to facilitate participation by IMLS digitization projects.
- Remote assistance with implementation of OAI compliant metadata provider services at IMLS grantee sites. May include assistance with upgrading and transformation of metadata as well as technical implementation of OAI protocol.
- Implement item-level metadata harvesting and repository service. Define terms and conditions, including required metadata and technical pre-requisites, for IMLS grantee participation in item-level metadata repository.
- Ongoing testing and refinement of item-level metadata repository service based on outcome of research investigations and feedback solicited from IMLS, participating projects, and selected end-user groups.

Research Investigations. As Lagoze and Fielding suggest,¹⁵ traditional library collection functions that attend to user-based criteria for selection are key to the success of distributed digital collection services. Criteria can be relatively straightforward where the range of content is narrow and the user base is homogeneous, as in the case of collections of research reports for computer scientists. Traditional collection problems of libraries and museums, however, take on new complexities in the digital environment. It has always been expensive and difficult to build heterogeneous collections that support the interests of diverse user communities, and this legacy problem stands as one of the greatest challenges for distributed digital collections. Our research agenda is devoted to investigating how resource developers can best represent collections and items to meet the requirements of divergent service providers and user communities.

The new multimedia content and tools for retrieval and manipulation added to digital libraries as they evolve complicate the process of adapting systems to various user populations. But, through research and development, important advances have been made in design for heterogeneity of both content and audience. For example, the Perseus project has continued to provide extensive classics resources suitable both for school children and advanced scholars,¹⁶ and the Berkeley Digital Library for watershed planning accommodates the needs of landowners, environmentalists, farmers, scientists, engineers, and citizens.¹⁷ A member of our team is currently engaged in a project to build a digital library for biodiversity survey fieldwork to be used by elementary students, amateur adult volunteers, and professional botanists.¹⁸ It is essential that we continue to design for this level of customization as we open up digital repositories for wider discovery and reuse.

As we gain in interoperability, we do not want to lose advances that have been made in adaptation and

¹⁵ Carl Lagoze et al., (1998). Above.

¹⁶ Crane, G. (1996). Building a digital library: The Perseus Project as a case study in the humanities. In Edward Fox, Gary Marchionini (Eds.) *Proceedings of the 1st ACM Conference on Digital Libraries* (pp. 3-10). New York: ACM.

¹⁷ Lisa R. Schiff, Nancy A. Van House, and Mark H. Butler (1997). Understanding Complex Information Environments: a Social Analysis of Watershed Planning. *Digital Libraries '97: Proceedings of the ACM Digital Libraries Conference*, Philadelphia, PA, July 23-26, 1997; p. 161-186.

¹⁸ An Internet Environment for BioDiversity Survey Collaboration and Verification. National Science Foundation, Information Technology Research/Information Management. P.I.: Bryan Heidorn, with Co-PIs: Carole L. Palmer, Michael Jeffords, and Marilyn Lisowski., January 2002-December 2004.

access for communities of users. Variations in metadata standards reflect the variant roles of digital objects and the different aims and practices of resource developers and their constituent user communities. For example, in the case of visual resources, Greenberg has shown that administration functions are well supported by EAD, but REACH elements are superior for supporting discovery.¹⁹ The TEI header represents attributes of texts recognized by scholars²⁰, and the GEM standard addresses the needs of educators.²¹ At this point in time, we have no systematic way to judge the value of these schemas from the perspectives of the broad range of resource developers, and we know little about how effective the application of the schemas is in providing the access and functionality required by potential service providers and user communities.

Recent empirical studies of scholarly information use conducted by one member of our team demonstrated the need for libraries to do much more to assemble information resources in a way that allows scholars to search across many different kinds of databases, archives, and collections.²² Through interviews and document analysis techniques the study specified how scholars identify and locate sources, the attributes they attach to sources, and why those attributes are of value to them. One important dimension of value is the role that the source plays in the scholarly process. A particular item can play different roles in different kinds of scholarly work, and some documents host multiple types of data or evidence. Moreover, under certain circumstances a particular source may be used in different ways within a single project. These findings raise important questions about how materials might be more richly represented or encoded to reflect their potential roles and the unexpected ways that collections of like content might be determined and assembled by scholars. Our research will take a similar approach to examining how well collection-level and item-level metadata reflect what communities value and look for when searching for and aggregating resources.

We believe a strength of this proposal is the mix of project personnel. Our approach to the project will combine the pragmatics of attempting to solve the problem of the particular with the exploration of research issues that allow an abstraction and generalization of the findings to the very large set of similar cases needing to combine and use metadata. The activities to be undertaken will allow us to analyze and experiment with a variety of approaches and issues that are fundamental to the effective production, refinement, combination and use of metadata in order to fulfill end users' real needs. We outline a sample of these larger research questions below.

Evaluate usefulness and appropriateness of the "Framework of Guidance for Building Good Digital Collections"²³ There is a growing research interest in the process of developing standards, particularly data standards to support interoperability, including but not limited to metadata schemas. We will identify key issues in this area, especially concerns of standards negotiation and evolution. Although it might be administratively convenient to impose a single standard on all affected parties, it is rarely possible. Thus mechanisms must be developed to support incremental adoption, and intermediate translation features. In essence, support for standards evolution. We will provide a case analysis of the issues encountered within this project, accounting for the wide scale, scope, medium, and content coverage of IMLS projects and submit a report recommending any desirable extensions and/or additions to the Framework.

¹⁹ Greenberg, Jane. (2001). "A Quantitative Categorical Analysis of Metadata Elements in Image-Applicable Metadata Schemas." *Journal of the American Society for Information Science and Technology* 52 (11): 917-924.

²⁰ See Text Encoding Initiative. The XML Version of the TEI Guidelines.

<http://www.hcu.ox.ac.uk/TEI/P4X/HD.html#HD7>.

²¹ See GEM project site. Creating GEM Metadata for Database Collections.

<http://www.geminfo.org/Workbench/CreatingGemMetadataForDBs.html>

²² Palmer, Carole L., and Laura Neumann. (Forthcoming). "Information Use Studies for Digital Library Development: The Case of Interdisciplinary Humanities Scholars." *Computers and the Humanities*.

²³ Priscilla Caplan et al., 2001. "A Framework of Guidance for Building Good Digital Collections." Available:

<http://www.imls.gov/pubs/forumframework.htm>

Delineate critical metadata functions and levels of granularity for different user communities. The first step in this work will be to identify a substantial base of IMLS metadata providers that represent a wide range of content and metadata approaches. Log and survey techniques will then be used to monitor online interactions with the repositories and assess how they are being accessed and used. We will target service-oriented sites, such as museums, state library based projects, historical societies, cultural heritage education centers, and scholarly encoding initiatives, for more in-depth case study of their operations, requirements, and the information practices of their user constituencies. Our analysis will focus on the aspects of service provision and use that are dependent on metadata schemas. Building on Greenberg's²⁴ categorical analysis of image-applicable metadata schemas, we will determine how elements within the schemas support discovery, use, administration, and authentication functions relative to the needs and uses of different providers and users. We expect that our research will increase the field's understanding of domain-specific metadata functions and the levels of granularity needed for different functions.

Develop user-centric tools and standards to support cross-collection searching by multiple communities. IMLS grantees serve a variety of user communities, with a corresponding range of needs, expectations, and expertise. Prior work has characterized many of these user groups (e.g., the survey of informational needs of visitors to museum Web sites by Victoria Kravchyna and Sam Hastings²⁵). While the focus of this project is on the implementation of a collection registry and metadata provider services, the development of tools or sets of new functionalities is pointless if the intended users of the system cannot understand how to use them, cannot see how the features benefit them, or do not believe that it is worth their while to bother investigating the system. To quote the cliché, “for the user the interface *is* the system”. We will explore the issues of developing an interface that optimizes learnability and efficient use. This will involve iterative design and formative evaluation. An example of the challenge is the issue of how or whether to make clear to the user the great heterogeneity of the multiple collections and their various different metadata standards. Do we even wish to provide a seemingly seamless interface to a repository that really is not seamless at all? We are not merely proposing a technology but the incorporation of a sociotechnical system for information access and use. Tools and standards to support cross-collection searching and metadata harvesting must integrate with the multiple organizations involved. An excellent tool that is in some way unacceptable to the stakeholders will fail. Thus we need to take into account the wider organizational issues in the design and deployment process. This will be done by observing, reflecting on, and analyzing the issues that arise throughout the project in order to produce recommendations for effective sociotechnical design.

Management Plan

The Principal Investigator, together with Co-Principal Investigators, shall provide the overall direction of the project. Principal Investigator Tim Cole is currently leading an OAI Metadata Harvesting project. Co-Principal Investigators and the project consultant have all been involved in successful digital library projects for funding agencies including IMLS and the National Science Foundation. Nuala Bennett in particular has served as Project Coordinator for the IMLS funded DCHC and TDC projects described above. Grainger Library will host the collection registry and metadata repository services created as part of this project. Grainger Library currently hosts both the Illinois OAI Metadata Harvesting project the TDC project. Grainger Library was previously home for both the Illinois Digital Library Initiative / D-Lib Test Suite test bed project and the DCHC project. The Schedule of Completion provides the details of how the work of the project will proceed.

²⁴ Greenberg, Jane., (2001). Above.

²⁵ V. Kravchyna and S. K. Hastings. 2002. “Informational Value of Museum Websites,” *First Monday*, vol. 7 (2). Available: http://www.firstmonday.org/issues/issue7_2/kravchyna/index.html

Budget

The project budget aims to accommodate the cost of service development, deployment, and maintenance, the cost of working with participating IMLS grantees one-on-one as necessary to obtain collection-level metadata and implement OAI metadata provider services, and the cost of research investigations outlined above. The availability of extensive existing expertise and infrastructure for OAI metadata harvesting will allow us to spend relatively less of our project budget on service development and deployment while still implementing high-quality and robust collection registry and metadata repository services. Approximately thirty percent of the requested project budget will go to researchers in the Graduate School of Library and Information Science for research investigations as described above.

Contributions

The University Library will contribute salary costs and associated fringe benefits for all time spent by the PI and Library co-PIs and consultant on this project. In addition, though no value is explicitly shown in the detailed budget for this, the Grainger Engineering Library and Information Center will host collection registry and metadata repository services on Grainger Library servers and provide necessary network infrastructure. Support and infrastructure contributed by the Grainger Library will include the use of workstations (over and above the initial workstation purchased for the Project Coordinator at the start of the project), all workstation software, maintenance of workstation hardware and software, backup and anti-virus services, and high speed, high bandwidth access to building and campus networks and the Internet. Grainger also will contribute server software and hardware (other than disk drives) required to implement, develop, and test collection registry and metadata repository services. IMLS is asked to provide necessary disk drive space for installation into Grainger server hardware. The Graduate School of Library and Information Science (GSLIS) will also make infrastructure contributions. The GSLIS Information Systems Research Laboratory (ISRL) maintains a technical infrastructure designed to facilitate information systems related projects housed within GSLIS. The infrastructure includes production-quality e-mail, web, and file/print servers to facilitate day-to-day tasks. In addition, several high-end web/computational servers are maintained to provide a general experimental infrastructure for ISRL researchers. Shared file space and authentication/authorization mechanisms (using Lightweight Directory Access Protocol) facilitates migration between experimental and production servers. Currently, there is over a half terabyte of data storage available between these systems. A 14 tape DLT jukebox provides regular backups for the data. The systems are on a 100 Mb/s Ethernet using Category 6 wiring. They receive a high-speed Internet through the University of Illinois. The configuration of the servers and backup facilities for the laboratory are fully described at <http://www.isrl.uiuc.edu/systems/resources.html>. Office space at GSLIS will be provided for a project Research Assistant.

Personnel

Principal Investigator Timothy W. Cole is the PI of the Mellon Foundation funded Illinois OAI Metadata Harvesting project and is Mathematics Librarian and associate professor of library administration at the University of Illinois at Urbana-Champaign. Cole was Co-PI of the Illinois Digital Library Initiative and D-Lib Test Suite projects. Cole is a current member of the OAI PMH Technical Committee, and was a member of the IMLS Digital Library Forum. He participated in drafting the “Framework of Guidance for Building Good Digital Collections.”²⁶ He is an expert in XML and related technologies such as XSLT and in the use of metadata schemas including the qualified and unqualified Dublin Core schemas.

Co-Principal Investigator Michael Twidale is associate professor at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. His research includes the investigation of contextualized systems evaluation and interface design to support collaborative learning

²⁶ Priscilla Caplan et al., (2001). Above.

and working. Twidale received a NCSA/UIUC Faculty Fellows Program award in support of his work with “Cyberdocents: an exploration of education and guidance in and around museums.”

Co-Principal Investigator Carole L. Palmer is assistant professor at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. Her background is in academic librarianship, and she teaches in the areas of information seeking and use and current trends in collections and information services. Her research examines how well existing information structures and tools support research and problem solving. She specializes in qualitative studies of how people find and use information and the barriers that deter this process. She is currently engaged in projects to develop digital libraries and knowledge discovery systems that support and promote diverse research collaborations. All of her work is aimed at development of information environments that are responsive to the practices of user communities, with a particular focus on improving information systems and services for interdisciplinary inquiry.

Co-Principal Investigator Nuala A. Bennett is Interim Coordinator of Digital Imaging and Media Technology Initiatives, visiting assistant professor of library administration, and Project Coordinator for the IMLS-funded Teaching with Digital Content project at the University of Illinois at Urbana-Champaign. She brings knowledge and expertise working with museums, libraries, archives, K-12 educators, and College of Education faculty. Her previous experience includes Research Information Specialist with the National Center for Supercomputing Applications and Research Programmer and Project Coordinator for medical informatics projects with the Community Architectures for Network Information Systems Laboratory at the Graduate School of Library and Information Science.

Co-Principal Investigator William H. Mischo is the Engineering Librarian and professor of library administration at the University of Illinois at Urbana-Champaign. Mischo was the PI of the Illinois Digital Library Initiative and D-Lib Test Suite projects and expert in XML, reference linking systems, simultaneous search systems, and end-user interface design for library applications.

Project consultant Beth Sandore is Associate University Librarian for Information Technology Planning and Policy, and Professor at the University of Illinois at Urbana-Champaign Library. Sandore has had extensive project management and program development experience, having headed up previous IMLS National Leadership Grants, and grants from the National Science Foundation, as well as from private sources such as the Intel Corporation. She has served in an advisory capacity for a number of groups on imaging and technology evaluation projects, including the U. S. Department of Education, the Getty Information Institute, the Andrew Mellon Foundation, and the Oregon Historical Society.

Project Evaluation

A desired outcome of this project is to demonstrate the achievability and usefulness of metadata sharing at both collection-level and item-level for the domain of IMLS National Leadership projects. Complete or nearly complete participation in the collection registry is an essential target and an appropriate standard against which this project should be measured. The lack of appropriate item-level metadata and/or technical infrastructure issues at some grantee sites will preclude complete participation in item-level metadata repository. Unanticipated issues of scale and performance also may limit achievable comprehensiveness of metadata repository. However, it's important to understand the qualitative difficulty or ease with which grantees are able to participate. Throughout the project, a sampling of participating and non-participating grantees will be surveyed, both to determine the degree of success in regard to the objective of complete participation and to identify issues and barriers that might be discouraging participation. A perceived lessening in barriers to and concerns about participation in item-level metadata sharing is another measurable intended outcome of this project. Finally, a desired outcome of this project is to better understand the scope and magnitude of potential benefits to end-users of collection registry and metadata repository services for the domain of IMLS projects. Selected user

groups will be targeted and data in the form of transaction logs and in-person and online surveys will be collected throughout the project to estimate magnitude and nature of potential end-user benefits.

Dissemination

The project will actively seek out appropriate electronic forums, such as listservs and online discussion lists, in which to alert the wider library and museum communities of the progress of this project. The participants are active in traditional academic publication arenas and fully intend to generate appropriate journal articles, conference papers, and avail themselves to professional presentations designed to disseminate the findings of the project in a timely manner. Since the participants backgrounds span across traditional library and museum experiences and include integral ties to the scholastic arena provided by the Illinois Graduate School of Library and Information Science, they are well positioned to take advantage of numerous forums in which this research can be exposed, exploited, and built upon.

Sustainability

The ultimate goal of creating an IMLS Collection Registry and Metadata Repository is to establish the usefulness and viability of sharing collection-level and item-level metadata in the context of digitization projects like those sponsored by IMLS. The work we propose will accomplish those goals and will lay the foundation for further exploitation of these technologies and approaches by IMLS and IMLS grantees. We anticipate this project will establish the desirability, value, and relatively low cost of a permanent collection-level registry for IMLS funded digitization projects. Because of its nature as a means of unifying and enhancing visibility of such projects, it is unlikely that any organization other than IMLS will pay to implement and maintain this service. While such a service cannot be maintained without some ongoing cost, most of the expense of such a service is in the establishment phase when decisions such as metadata schemas and interface features and functions must be defined, tested, and refined. These costs will be paid by the conclusion of this project. Consideration will be given to constructing and implementing our prototype collection registry service in a manner to insure easy and inexpensive portability to IMLS or their long-term designee for this service. All software, documentation, and practices developed for the collection registry service will be included in deliverables to IMLS. We will also report on the maintenance costs and issues associated with long-term continuation of this service.

We also anticipate that this project will establish the desirability and value of sharing item-level metadata by IMLS grantees. However, it is less certain at this time that the IMLS will need or want to maintain its own separate item-level metadata repository. Harvesting may be best left to other commercial or non-commercial entities or organizations such as large-scale digital library projects, commercial search engines, library catalog vendors, state libraries, or regional consortiums. In the metadata sharing envisioned by the OAI PMH, many such organizations may in fact co-exist, harvesting selected subsets of metadata as appropriate to their particular missions. NSDL, as representative of a large-scale digital library project, has already embraced the OAI PMH. State library initiatives such as the Find-It! Illinois service of the Illinois State Library are prime potential future consumers of metadata made available via OAI PMH. Library vendors and Internet search services also are expressing an interest in the potential of OAI PMH to support more efficient and comprehensive Web searching. With regard to item-level metadata sharing, a primary objective of the proposed work is to engender a commitment on the part of IMLS grantees and other cultural heritage institutions to implement and maintain metadata provider services so that metadata may be harvested by interested parties. This project will accomplish this objective by creating a range of prototype metadata provider tools and metadata transformation services that can be adapted across the whole spectrum of library and museum digitization activities. It will also document the costs and benefits of implementing and maintaining such services.