

**IMLS National Leadership Grant LG-02-02-0281**  
**"Proposal for IMLS Collection Registry and Metadata Repository"**

*This summary is part of the three-year interim project report for the IMLS Digital Collections & Content Project, summarizing major findings October 2002 through September 2005. Project is hosted at the University of Illinois at Urbana-Champaign. Project Director is Timothy W. Cole ([t-cole@uiuc.edu](mailto:t-cole@uiuc.edu)). Full report is available at <http://imlsdcc.grainger.uiuc.edu>. The material in this report is based upon work supported by the Institute of Museum and Library Services under IMLS National Leadership Grant Award No. LG-02-02-0281. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Institute of Museum and Library Services.*

**Digital Collections & Content Project:**  
**Three-Year Interim Report Executive Summary**

**Introduction**

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, collection-level description and the organization of content into collections remains highly relevant (Miller 2000). 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. Sharing item-level metadata within a collection or repository has the potential to enhance the discoverability of individual items. Long-term value and utility of digitized content is greatly enhanced through inclusion in a collection registry and, when appropriate to the nature of a collection, the implementation of item-level metadata sharing protocols.

During the first three years of this project, we designed and implemented a collection-level registry and item-level metadata repository service that aggregates information about digital collections and items of digital content created or developed with resources from Institute of Museum and Library Services (IMLS) National Leadership Grants. Design and implementation have been informed by concurrent research. This project has created opportunities for us to examine digital collection and descriptive metadata practices from a variety of projects with differing backgrounds and community traditions. Areas of special emphasis have included collection descriptions and the associated collection registry; item-level description and the associated metadata repository; and the application and use within the IMLS grantee community of the IMLS/NISO *Framework of Guidance for Building Good Digital Collections*. This executive summary discusses significant observations to date and describes preliminary project findings and recommendations for IMLS and the broader digital library (DL) community.

## General Findings & Recommendations

Work to date has surfaced the following findings and recommendations. Specific suggestions and recommendations to IMLS are highlighted and boxed:

- Virtual aggregations of digital resources require clear and complete collection development and collection selection policies. Selection on administrative or programmatic criteria alone lead to aggregations of limited scope and utility.
- Examination of the collection registry and item-level metadata repository reveals multiple areas of collection strength in digital content so far developed under the auspices of the NLG program.

**IMLS should exploit these collection strengths through continued/expanded collaboration with relevant existing nationally scoped initiatives (e.g., NSDL, GEM) and/or by extending collection selection criteria for the collection registry and metadata repository to encompass related resources (e.g., Library of Congress American Memory, published secondary resources and aggregations).**

- Both content provider and service provider share responsibility for the successful aggregation of digital resources. Content developers need to be concerned not just with their local project but also with the contribution they can make to larger initiatives. Centrally supported service providers must accept responsibility to normalize and adapt metadata for use in the context of services implemented.
- Some content management systems limit the capacity of content providers to share metadata in optimum ways.

**IMLS should encourage more collaboration between resource providers and vendors to improve application sharing features.**

- Many NLG projects emphasize digital content created for K-12 audience. Often the cultural heritage resources represented in these collections have broad appeal to many other audiences. This potential should not be overlooked or undervalued.
- Federation should not be one dimensional. Digital collections and collection components can be included to good purpose in multiple aggregations of varying size and scope, just as a digital artifact can be in many collections at once.
- The definition of digital collection is evolving as digital projects become more sophisticated.

**IMLS should encourage research to determine how collections identify themselves and how digital collection definitions are evolving.**

- Examination of collection registry transaction logs confirms the importance of subject searching (broadly defined), but also suggests that concept-only subject classification of collection registry records (e.g., using GEM subject headings) is insufficient.

**IMLS should encourage further research into the nature of subject searching in federated context and into additional vocabularies that support user needs.**

- Because digital projects are relatively new to the library world, the evolution of digital projects over extended periods is yet to be fully understood.
- While community best practices, guidelines, and standards provide useful targets for projects, these targets can be difficult to realize in practice, especially in the short term. To be most useful, emerging guidelines in this domain should be cognizant of the difficulties involved in implementing successful projects, and should show appropriate steps for successfully utilizing their recommendations.

### **Collection Description & Collection Registry**

Assessment of the collections and associated descriptions, including research focusing on the nature of collections and sub-collections, surfaced the following key findings:

- Few digital resource developers articulate formal selection/inclusion policies or guidelines. Existing collection development policies most often emphasize audience, geographical and temporal scope, original physical collection(s), sometimes preservation needs, value and significance of documents.
- A general sense among digital resource developers is that end users do not care about collections and are not assisted by collection description or orientation in resource design. Findings have shown that in the item-level metadata repository, both item and collection-level metadata are essential: item-level description supports retrieval of objects, while collection-level description represents uniqueness, authority, and context of objects.
- Concept of collection remains ambiguous, blurred with the notion of project.
- Although 75% of digital resource developers report division into sub-collections, only 2.4% list/mention sub-collections in collection description records. The collection-sub-collection relations are hard to define, many digital resource developers experience difficulty differentiating collections from each other and from their sub-collections.
- Audience is often more diverse than anticipated at initiation of a project; however, actual audiences are taken in consideration during a project when customizing collections.
- Current collection-level metadata schemas do not differentiate the properties of the collection as a whole from the properties of the items as individual collection members.

Project managers are allowed to edit their collection-level records online when the collection is added to the registry. Observing this process we found most resource developers enrich rather than refine their collection level records, with Audience (with a trend for widening), GEM subjects, Size, Frequency of Additions, and Geographic Coverage fields modified most often.

The project developed an interface for browsing and searching the collection registry, and conducted usability surveys in 2005. Users can currently browse the collection registry by GEM subject, temporal coverage, spatial coverage, title, grant project, and hosting institutions. The following are findings from usability studies of the collection registry interface:

- Although most users were unfamiliar with this form of aggregation, they demonstrated a quick learning curve for familiarity with the site and the services it provided.
- The collection development approach for this site (IMLS NLG recipients) was not a natural aggregation for the end user.

Analysis of the subject searches performed in the collection registry reveals the following:

- The broadly defined subject search (both controlled- and uncontrolled-vocabulary search with intent to find information on particular subject/topic/discipline/area) prevails and accounts for 70% of all searches made by users between February and September 2005. This number is significantly higher than reported by transaction log studies of online catalog use in the 1980s and 1990s. Such an increase can be explained by at least two reasons: 1) a general shift towards subject search due to exponential growth of publication further limiting user's ability to select the specific title or author to search for; 2) a conceptual difference between collection-level and item-level search, which implies a trend towards increased levels of subject search in federated collection registries compared to single collections.
- Being confined to *concepts* and ignoring other significant groups of subjects -- *objects, places, events, corporate bodies, persons* etc. -- GEM subject scheme appears incapable of representing subject scope and breadth of collections in the collection registry. Only 2.6% of the user searches made in the collection registry between February and September 2005 were semantically matched by GEM scheme. Art and Architecture Thesaurus does not provide a fully satisfactory alternative, being limited to *concepts* and *objects* and matching only 22.63% of user searches; LCSH demonstrates a rather high (71.3%) level of semantic match with user queries.

Investigations into the perceptions of possible use of the collection registry found that many resource developers are unsure of the role and value of federated resources for their institutions. Only 40% recognized potential benefits for reference and research services, few perceived the IMLS DCC collection registry as a helpful tool for end users. The registry is often viewed as a source of information on up-to-date practices for digital projects and grant funding trends.

### **Item-Level Description & Metadata Repository**

The metadata repository has allowed for examination of metadata schemas and associated item-level metadata from diverse institutions. As expected, we found variations in metadata standards and usage reflecting the variant roles of digital objects and the different aims and practices of resource developers and their constituent user communities. Major findings in this area:

- Scheme selection is influenced by the degree to which scheme was implemented / tested; use by peer or collaborating institutions; compatibility with local systems; and local familiarity with the scheme.

- About one-third of the projects utilize multiple schemas. MARC and DC formats are most used, with MARC usage gradually declining and DC usage gradually increasing. Limited application to date of more specialized schemes (EAD, VRA).

The project developed a workflow to normalize and adapt harvested metadata values and semantics (i.e., schema) for use in item-level search portal. This effort drew heavily on work done by both the NSDL and the Western States Digital Standards Group (available at <http://www.cdpheritage.org/index.cfm>). Concurrently interviews were conducted with project participants regarding their approach to sharing metadata and resulting in the following findings:

- Federation is rarely taken into consideration when designing digital projects.
- Primary concerns for federation include: 1) no one scheme can meet expectations and needs of all cultural heritage institutions; 2) dissatisfaction with sparseness of DC, which is most widely used, in part for OAI compliance; 3) emerging metadata quality concerns related to consistency, granularity, and integration.
- Many institutions are not well positioned to bear the cost of developing high-quality, sharable metadata.
- Problems can occur when the original context of the metadata is lost.

The project team analyzed harvested metadata in order to understand how best to optimize metadata for a shared environment. Results suggest the following findings and recommendations regarding metadata used in a shared context (see also Shreeves, Riley, and Milewicz 2006):

- Shareable metadata must accurately describe the resource.
- Shareable metadata must be coherent.
- Individual metadata records should not depend on the local context.
- Consistency of metadata can be more important than completeness.
- For best outcome, metadata should conform to nationally established standards.
- Good communication between content and service providers is critical.
- Richer schemes than Dublin Core should be exposed, if available.
- Content management systems create technical difficulties for sharing metadata.

Although the majority of IMLS NLG digital projects include item-level metadata, we found several barriers to implementing OAI-PMH. In particular:

- Technical infrastructure, whether computing or staff resources, is not available for implementation, ongoing support, or is in transition (delaying consideration of OAI).
- Proprietary system in use does not have OAI capability or the available OAI data provider is seriously flawed.
- An investment has already been made for another means of sharing metadata.

- Metadata is not in a shareable state.
- Metadata is too complex to be represented well in simple Dublin Core.
- Agreement of all partners in a collaborative project may be required to expose metadata.
- The collection of items is not yet public.
- Interoperability and OAI in particular is not a priority or is unfamiliar.

### **Report on the *Framework of Guidance for Building Good Digital Collections***

Of interest to our project from the outset has been the degree to which principles articulated in the IMLS/NISO *Framework of Guidance for Building Good Digital Collections* are being followed in the IMLS grantee community. Based on our experience and results to date we offered several recommendations to NISO to consider in developing future editions of the Framework. These recommendations are described in full in the attached sub-report. Major recommendations include:

- Include principles that span multiple of the entities (collection, object, metadata, projects). Combine and/or broaden existing principles tied to single entities which encourage good documentation, sustainability, measurements of usefulness, and descriptions of IP rights with basic entities. Consider new principles that cut across multiple entities.
- Elaborate existing principles and/or include additional principles in order to better address the emerging significance of collection-item and collection-collection relationships, and stress especially the value of describing collections in the context of other digital resources, i.e., encourage outward-looking description in addition to inward-looking views of the collection.
- Stress the obligations of DL service providers and their role as collaborators with data providers in enabling and facilitating delivery of DL services across distributed collections of content through the staged creation, normalization, remediation, and enrichment of metadata at multiple points in the metadata use cycle.

Our review of the *Framework* also suggested eight research opportunities IMLS may wish to encourage. These are described in the attached sub-report. Three of these eight research opportunities are highlighted here:

- **RESEARCH OPPORTUNITY 1: Explicitly encourage projects, research, and other work that will help operationalize the *Framework*.**

- **RESEARCH OPPORTUNITY 3: Encourage work that features further research regarding and/or exploitation of structured collection-level description, descriptive granularity, relationships between collections and items / other collections.**

- **RESEARCH OPPORTUNITY 5: Because the differences between use and usefulness are not clear to DL managers, encourage research informing and providing guidance for assessing digital collections and evaluating their usefulness.**

## Metadata Roundtable

In 2003 a metadata roundtable was initiated to bring together local and visiting practitioners, teachers and students to discuss and analyze the latest developments in metadata theory and practice. This model provides a unique opportunity to complement classroom learning with the practical expertise and experiences of practicing librarians and the research results and methodologies of graduate school faculty. Topics discussed at this roundtable have included metadata quality, interoperable metadata, and approaches for adding value to metadata (see attached report for additional information). Since its inception in the spring of 2003, 67 meetings have been conducted, and, due to the increasing popularity, meeting regularity has increased from twice a month to weekly.

## Conclusion

A desired outcome of this project was to demonstrate the achievability and usefulness of metadata sharing at both collection-level and item-level for the domain of IMLS National Leadership Grant projects. We believe that our project already has successfully met this goal through the establishment of the Collection Registry and Item-level Repository. Additionally, the team investigated issues surrounding collection identity, collection-level metadata, interoperable metadata, and barriers to OAI-PMH. This work will inform future metadata aggregation projects as well as individual projects hoping to make their metadata more useful to such projects.

Another desired outcome of this project was 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. Our usability studies have furthered our understanding of how end users approach aggregations, and informed ongoing development of the search and browse interfaces.

Another notable outcome has been the success of the collaboration between the University Library and the Graduate School of Library and Information Science (GSLIS). This collaboration has strengthened both practical implementations and research results. Less anticipated, this collaboration also has encouraged collaborative work beyond the immediate scope of the DCC project and has provided opportunities for enhanced interactions between Library and GSLIS faculty and GSLIS students. Additionally, the metadata roundtable has created a new model for broader sharing, collaboration, and dissemination of advances in the state-of-the-art metadata concepts for all interested individuals across the university.

Our findings from the first three years of the IMLS Digital Collections and Content project will continue to inform the development of the project during the remainder of the grant period.

## References

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

Shreeves, Sarah L., Jenn Riley, Liz Milewicz (2006). "Moving Towards Sharable Metadata." In *First Monday*, 11 (8), Available:  
<[http://firstmonday.org/issues/issue11\\_8/shreeves/index.html](http://firstmonday.org/issues/issue11_8/shreeves/index.html)>