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

This white paper 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). 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.

Preliminary Recommendations for Shareable Metadata Best Practices White Paper

Amy Jackson

Digital Collections and Content Project
Grainger Engineering Library
University of Illinois at Urbana-Champaign

October 16, 2006

Introduction

A primary goal of the IMLS Digital Collections and Content project has been to encourage sharing of metadata and participation in aggregating projects. Benefits realized through these activities include increased visibility of the collection, an environment of complimentary resources, and enhanced discoverability of individual resources. In order for these benefits to be achieved, metadata from each repository must be interoperable with existing metadata in the aggregation. Although metadata may be quality metadata in the local context, the records must have certain characteristics in order to be relevant in the aggregated environment.

While the responsible for producing quality metadata records resides with the data provider, the data providers should not be responsible for anticipating all future uses and contexts surrounding each individual metadata record. Service providers expect to minimally process the acquired data in order to ensure consistency across metadata harvested from different sources. This processing is minimal; each individual record is not examined, and the service provider does not enhance existing information. The data provider should recognize that the service provider does not have an equivalent level of expertise regarding the subject matter, and should strive to provide records in a complete and interoperable format.

In an ideal environment, all metadata would be interoperable. In reality, this rarely happens. Several barriers to interoperable metadata have been encountered during the project including:

- No one scheme can meet expectations and needs of different cultural heritage institutions;
- Dissatisfaction with sparseness of Dublin Core, which is most widely used, in part for OAI compliance;
- Emerging metadata quality concerns related to consistency, granularity, and integration;
- Problems when the original context of the metadata is lost;
- Federation is rarely taken into consideration when designing digital projects;
- Many institutions are not positioned to bear the cost of developing high-quality, sharable metadata.

Even though most institutions are faced with at least some of these barriers, attempts should be made to overcome these challenges in order to achieve benefits from contributing metadata to a shared environment. To this end, we have developed preliminary recommendations for shareable metadata best practices (see also Shreeves, Riley, and Milewicz 2006; and NSDL OAI Best Practices for Shareable Metadata). Following these recommended standards will ensure that metadata created in a local environment will continue to be meaningful in a broader context.

Shareable metadata records should not be dependant on the local context.
Although many institutions create metadata appropriate for their local context, in order for metadata to be truly shareable, it must be relevant in many different contexts. Associations that are assumed inside of a specific collection loose their relevance when the metadata is aggregated outside of its native context. Robin Wendler discusses this contextual concept in her 'on a horse' problem, when the collection is about a specific individual, but the individual is not mentioned at the item-level (Wendler 2004).

Shareable metadata accurately describes the resource.
In order for an individual metadata record to stand on its own, the end user should be able to describe the resource described in the record. Important keywords should be included describing both the format and the subject(s).

The content of a metadata record should be optimized for sharing.
In the shared environment certain characteristics of metadata records ensure their optimal use. At a minimal level, completed title and identifier fields are highly recommended for usefulness. The identifier should link directly to the resource being described, and the primary link should be to the item with its contextual material. Additionally, metadata should be exposed at the most granular level appropriate for the described resources. Granularity varies across different types of collections, and resources maybe described at the item level, the collection level, or a sub-collection level. If each photograph in a collection is described individually, the records for each photograph should be exposed. On the other hand, metadata for each individual page of a book is too small of a level. Metadata for the entire book should be exposed in place of each individual page. Optimal metadata records also do not contain information useful for only the local institution; administrative and preservation metadata is generally irrelevant in the

aggregated context. Because not all metadata is necessary outside of the local context, the institution may hold a master record and expose only appropriate elements for harvesting.

Consistency of metadata is more important than completeness.
When metadata elements are handled consistently across a collection, service providers can easily manipulate harvested metadata to fit within locally defined fields. In order for these transformations to be successful, the acquired metadata must be consistent semantically and syntactically. Examples include date and coverage fields, delineation of multiple subjects, use of controlled vocabulary, and sub-collection affiliation.

Metadata should conform to nationally established standards.
Adhering to nationally recognized controlled vocabularies, encoding standards, and published metadata schemes ensures that all metadata in an aggregated environment will be interoperable.

Data providers should communicate with service providers regarding their metadata usage.
Both service providers and data providers benefit when metadata usage is clearly explained. With strong communication, the service provider is able to filter and normalize the data while retaining the rich information included by the data provider. Information helpful for a service provider includes vocabulary and content standard usage, information regarding how often records are added and edited, and problems the provider may encounter while attempting to harvest the records. The data provider benefits from this communication by learning about metadata practices used in similar projects and how to create useful metadata for the user community.

Richer schemes then Dublin Core should be exposed, if available.
OAI-PMH requires that all metadata be exposed as Dublin Core records. However, many institutions follow or have developed richer schemes appropriate for describing their resources, and these schemes may also be exposed for metadata harvesting. OAI-PMH stipulates that all exposed schemas must have a valid and published XML schema. If the richer schemes are harvested, data providers can map to locally defined fields and provide the richer information for their end users. If crosswalking is performed at the local institution, always map from a richer scheme to a simpler scheme. Transformations in the opposite direction will loose valuable properties of the richer scheme.

Use appropriate tools to optimize the interoperability of metadata records.
Metadata tools include schemas, controlled vocabularies, content management systems, and community standards. Each of these tools and their roles and expectations in the intended user community, should be examined in order to provide the highest level of usefulness and interoperability. Users of content management systems incapable of exposing metadata should work with vendors to ensure future releases have these capabilities.

The attached presentation was developed as a workshop introducing the range and complexity of issues around metadata in the digital library environment.

References

NSDL OAI Best Practices for Shareable Metadata. Available: <http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl?PublicTOC>

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>

Wendler, Robin (2004). "The eye of the beholder: Challenges of image description and access at Harvard," In: Diane I. Hillmann and Elaine Westbrooks (editors). *Metadata in Practice*. Chicago: ALA Editions, pp. 51–69.