

Chapter 9 Supplement

Metadata Interoperability, Shareability, and Quality

Questions for Review, Study, or Discussion

1. This chapter deals with a set of interrelated topics all having to do with the usability of local metadata outside of its original closed-system context. What are those topics? What is *interoperability*?
2. What questions might any current digital collection implementer ask about the short and long-term viability of their metadata? How does this relate to designing local metadata schemes and creating current metadata?
3. In addition to its original context, which is often designed for a specific collection, cultural heritage metadata is also often shared in what three ways noted in the text?
4. Once local metadata is harvested and aggregated into an external repository, what is lost?
5. What is the Open Archives Initiative (OAI)? What is the OAI Protocol for Metadata Harvesting (OAI-PMH)? What is the difference between, and role of, data providers and service provider in OAI harvesting? What is OAIster?
6. What is the difference between aggregating and federated searching of metadata?
7. What metadata formats or schemes does the OAI Protocol support? What does it always require as a minimum for a basic level of guaranteed interoperability?
8. What are metadata *mapping* and *crosswalks*? How are these two terms used in relation to each other? What are some examples of metadata mappings or crosswalks in textual document form and in machine-actionable form?
9. Where do you see potential data loss when mapping from collection-specific local elements to simple Dublin Core, and from Qualified Dublin Core to Simple Dublin Core?
10. What are some common problems in semantic element mapping? Can you give concrete examples of these and other issues when mapping between Dublin Core and MODS, in both directions?
11. What does it mean to speak of Dublin Core as a “pidgin” or “switching” language?
12. What does this chapter mean by the terms *conversion*, *processing*, *normalization*, *manipulation*, and *massaging* of metadata? (Note: these terms overlap in meaning to various extents as used in this chapter.) In what contexts are these processes performed?
13. Look at the example of metadata harvesting, processing, and aggregating in Section 9.7 on pages 237-240. What changes do you see in the metadata from its original format to its two harvested and processed formats? Where do you see changes in the metadata properties or elements? Where do you see changes in the position and relationship among the metadata values?
14. What are some characteristics of metadata quality? What are some benefits of creating shareable metadata?
15. What are some methods of assessing metadata quality? How can Excel or other basic spreadsheet software be used for this purpose? What kinds of quality issues can be revealed by sorting on various columns after metadata has been exported into a spreadsheet?
16. The chapter concludes with five ways to improve metadata quality and interoperability. What are they? Do you agree with them and/or see how and why they would do so?
17. Do you have any personal experience with any of the topics covered in this chapter?

Recommended Readings and Resources for Reference or Further Study

Readings

- Bruce, Thomas R., and Diane I. Hillmann. 2004. "The Continuum of METADATA Quality: Defining, Expressing, Exploiting." In *Metadata in Practice*, edited by Diane Hillmann and Elaine Westbrooks, 238–256. Chicago: ALA.
- *Cataloging & Classification Quarterly* 46, no. 1. Special issue on Bibliographic Database Quality. Table of Contents: <http://www.catalogingandclassificationquarterly.com/ccq46nr1.html>.
- DLF/NSDL Working Group on OAI PMH Best Practices. 2007. "Best Practices for OAI PMH Data Provider Implementations and Shareable Metadata." Digital Library Federation. <http://www.diglib.org/pubs/dlf108.pdf>.
- Dushay, Naomi, and Diane Hillmann. 2003. "Analyzing Metadata for Effective Use and Re-Use." In DC-2003—*Proceedings of the International DCMI Metadata Conference and Workshop*, September 28–October 2, 2003, Seattle, Washington, USA. <http://dcpapers.dublincore.org/ojs/pubs/article/view/744/740>.
- Han, Myung-Ja, Christine Cho, Timothy W. Cole, and Amy S. Jackson. 2009. "Metadata for Special Collections in CONTENTdm: How to Improve Interoperability of Unique Fields Through OAI-PMH." *Journal of Library Metadata* 9, no. 3/4: 213–238.
- Hutt, Arwen, and Jenn Riley. 2005. "Semantics and Syntax of Dublin Core Usage in Open Archives Initiative Data Providers of Cultural Heritage Materials." In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, Denver, CO, USA, June 7–11, 2005, 262–270. New York: ACM Press.
- Jackson, Amy S., Myung-Ja Han, Kurt Groetsch, Megan Mustafoff, and Timothy W. Cole. 2008. "Dublin Core Metadata Harvested Through OAI-PMH." *Journal of Library Metadata* 8, no. 1: 5–21.
- Riley, Jenn, John Chapman, Sarah Shreeves, Laura Akerman, and William Landis. 2008. "Promoting Shareability: Metadata Activities of the DLF Aquifer Initiative." *Journal of Library Metadata* 8, no. 3: 221–248.
- Shreeves, Sarah L. 2005. "The Open Archives Initiative Protocol for Metadata Harvesting." In *Technology for the Rest of Us: A Primer on Computer Technologies for the Low-Tech Librarian*, edited by Nancy Courtney, 85–108. Westport, CT: Libraries Unlimited.
- Shreeves, Sarah L., Jenn Riley, and Liz Milewicz. 2006. "Moving Towards Shareable Metadata." *First Monday* 11, no. 8. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1386/1304>.
- St. Pierre, Margaret, and William P. LaPlant. 1998. "Issues in Crosswalking: Content Metadata Standards." NISO Standards White Paper. October 15. http://www.niso.org/publications/white_papers/crosswalk/.
- Woodley, Mary. 2008. "Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information." In *Introduction to Metadata*. 2nd ed., v. 3.0, edited by Murtha Baca. Los Angeles: Getty Research Institute. http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html.

Technical Documents

- DLF (Digital Library Federation). 2007. "Best Practices for OAI Data Provider Implementations and Shareable Metadata." Last modified July 10. http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/Main_Page.

- DLF (Digital Library Federation). 2008. "Introduction to Best Practices for Shareable Metadata." Last modified January 17.
<http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/IntroductionMetadataContent>.
- OAI (Open Archives Initiative). 2008. "OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting) Protocol Version 2.0 of 2002-06-14." Document Version 2008-12-07T20:42:00Z.
<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>.

Mappings/Crosswalks

- Library of Congress. MARC Standards: MARC Mappings. <http://www.loc.gov/marc/marcdocz.html>.
- Library of Congress. MODS Conversions: MODS Mappings.
<http://www.loc.gov/standards/mods/mods-conversions.html>.
- "Metadata Standards Crosswalk." 2009. Research at the Getty. Last modified June 9.
www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html.
- MIT Libraries. "Metadata Mappings (Crosswalks)."
<http://libguides.mit.edu/content.php?pid=158753&sid=1343490>.

Exercises

Recommended Exercises

1. Evaluate one or two metadata records in terms of their overall quality and shareability, and the extent to which they can stand on their own to make the digital resources they represent both findable and interpretable after mapping to simple Dublin Core and taken out of their original context. What would need to be changed or specified in the local metadata scheme or application profile in order to make the metadata sharable and meaningful outside of its original context?
2. Take a MODS record and map it to simple Dublin Core following the MODS mapping table from the MODS Website, also provided in this chapter. Where are there ambiguities and challenges in mapping data from one element set to the other? Where is there loss of richness, context, and/or meaning? What potential functionality is lost?
3. Take a set of metadata records in a spreadsheet and using column sorting to assess various aspects of metadata quality, such as completeness and consistency.
4. Take a set of metadata / database records for resources not following Dublin Core, MODS, or any other established cultural heritage metadata standard, if you have access to such records, and decide what you need to do to map the elements or fields into a Dublin Core, MODS, or VRA environment. What kinds of processing, conversion, manipulation, and normalization needs to be performed on the data values in order to make the consistent and to conform to one or more syntax and vocabulary encoding schemes?

See also the examples in the companion website supplementary document titled "[Metadata Application Profiles, Records, Functionality, and Quality Examples](#)."

Suggestions for Instructors

5. For Exercise 1 above, provide students with at least two metadata records, one that stands well on its own as simple Dublin Core in an aggregated context, and one that does not. Have students evaluate the records and answer the questions provided above. Consider also providing tabular summary application profiles or data dictionaries for assessment. This may be done as a written and/or discussion exercises.

6. For Exercise 2, provide students with a MODS record for mapping and have them share and discuss their results.
7. For Exercise 3, provide a spreadsheet of 20-30 records that include elements with missing and incorrect data values, and have students perform sorting to assess metadata quality. This could be presented as a scenario of a local metadata coordinator assessing records created for a local digital collection, or as a scenario of a member of a consortium or other aggregator assessing metadata records submitted for harvesting and inclusion in their repository. What processing would they need to perform before incorporating the records into their repository?
8. Use an imaginary scenario, but one that nonetheless mirrors actual practice, in which an institution is ingesting nonstandard metadata from one or more sources created according to some person's or organization's own invented scheme. A realistic scenario would be a university library taking over a collection of digital resources from a university department and working with the metadata created in an Access database, in which the creators were not dealing with any kind of standardized scheme such as Dublin Core, and did not use standard methods for recording dates, names, subject terms, and the like. Have the students create a mapping table from the nonstandard scheme to a Dublin Core or MODS based scheme. Have them assess what kinds of programs would need to be written to convert, process, and normalize nonstandard data values to standard ones. You may provide some real or invented examples of nonstandard metadata or ask students to invent some. A variation on this exercise is to have each student create her or his own metadata scheme using her or his own collection-specific element names, and perhaps various practices for recording uncertain dates and the like, and have students work in pairs to decide what they need to do in order to make their schemes work together.

See also the examples in the companion website supplementary document titled "[Metadata Application Profiles, Records, Functionality, and Quality Examples.](#)"