

Metadata Systems 2004-2006

Introduction

The topic of metadata is of great importance to libraries and librarianship at present. Articles have been written on the importance of cataloger involvement in digital library projects and the skills they bring to metadata creation (DeZelar-Tiedman 2004). The literature has shifted away from whether or not MARC is dead to how MARC operates as one of the many metadata options (Eden 2004). The confluence of various discrete areas (cataloging, digital libraries, e-journals, open-URL resolvers, etc.) and their integration, will determine the success of libraries in serving the users' needs now and in the future. How libraries manage their metadata may have a direct correlation to their relevancy in the future.

In order to narrow a hopelessly broad topic, this essay focuses on information published within the library and information science literature between 2004 and the first half of 2006. The topic of metadata is by no means exclusive to the library and information science community. Metadata has been written extensively about in computer science and business.

This brief essay is not intended to be comprehensive. Rather, it is meant to build on the earlier metadata systems essays written in 1995-2000 and 2000-2001 respectively

(<http://www.ala.org/ALCTSTemplate.cfm?Section=Sections2&template=/ContentManagement/ContentDisplay.cfm&ContentID=130413>). The essay will highlight some of the current literature on metadata.

In this third update of suggested research topics and essay methodologies on metadata, it is possible to see the literature moving beyond detailing the various metadata schemas and definitions of metadata into topics such as interoperability, non-descriptive metadata (e.g., administrative/rights metadata, preservation metadata, etc.), and whether or not various metadata schema are being applied in ways that assist users in finding what they are seeking. This shift is the hallmark of a maturing topic.

Beyond traditionally indexed publications (both print and electronic) exists a wealth of grey literature on metadata. Conference papers and presentations, while valuable, are not included in this review. Finally, it is worth noting that a 2005 special issue of *Cataloging & Classification Quarterly* (v. 40, issue 3/4) is devoted entirely to metadata. MARC and metadata were also featured in a 2004 double issue of *Library Hi Tech* (v. 22, issues 1-2).

Overview of metadata schema and standards

Descriptive metadata has come of age. Articles on various metadata schemas/standards and their related topics (Dublin Core, XML, RDF, METS, MODS, MADS) are abundant. Many articles include specific examples of their

application (Cundiff 2004, El-Sherbini and Klim 2004, Coyle 2005, Hillmann and Westbrook 2004, McCallum 2004, Smiraglia 2005). Others frame the selection of metadata schema a bit more broadly, seeking a single schema to cover all the contents of a repository, “to convert the disparate metadata to a consistent vendor neutral format ...” (Goldsmith and Knudson 2006). The later is a departure from those who chose a principal schema and augment from other schemas as necessary.

Intner, Lazinger, and Weihs (2006) provide a solid overview of different metadata schemas and how each schema relates to Functional Requirements for Bibliographic Records (FRBR). Coyle (2004) suggests that greater flexibility in record structure is needed and that FRBR, combined with description, discovery, and promotion metadata, moves library systems beyond MARC. Tennant (2004) proposes various means of creating diversity within local systems to handle non-MARC metadata (crosswalks, merging, and system migration).

Howarth (2005) discusses the 2004 work to date by the International Federation of Library Associations (IFLA) Cataloguing Section Working Group on the Use of Metadata Schema. Chohey (2005) looks at the metadata expertise required to plan and implement a digital repository through the eyes of a cataloger. In another article, Howarth (2004) provides a list of thoughtful questions to consider when developing a subject gateway. Many of the questions are applicable to any digital project.

With so many metadata schema options, how does one compare what is available? Polydoratou (2006) describes Germany’s MetaForm, a European metadata registry database that provides descriptions and usage information on various metadata schemas, especially Dublin Core. Wagner and Weibel (2005) discuss the Dublin Core metadata registry and how it benefits communities wishing to employ Dublin Core in some fashion.

Robertson (2005) argues that different types of repositories (digital libraries, institutional repositories, subject repositories, learning object repositories) have different needs in the quality of metadata required as each has diverse standards and purposes. One size does not fit all. Robertson suggests the nature of resource should dictate the amount of metadata created. How are institutional repositories being integrated with traditional and non-traditional resources?

Interoperability and shareable metadata

As witnessed before with the development and refinement of MARC, interoperability is crucial. Metadata interoperability can be approached at a number of different levels. Metadata at the schema (application neutral), record (mapping), and repository levels (connecting data strings with given elements across multiple sources) all achieve interoperability through different means (Chan and Zeng 2006; Zeng and Chan 2006). Chan (2005) lists various definitions of interoperability and details the seven most common methods used

to create interoperability in descriptive metadata (uniform standard, application profiling, derivation, crosswalk/mapping, switching schema, lingua franca, and metadata framework/container). Weibel (2005) muses on the challenges of collaboration and consensus in the digital age with particular attention to interoperability and some of the barriers including the “not invented here syndrome,” parallel metadata developments, differing functional requirements, and the costs associated with collaboration.

Medeiros (2006) states that interoperability in terms of cross-searching and information retrieval often conflicts with local information communities needs. Medeiros highlights the findings from the draft report entitled *Best Practices for Shareable Metadata* (2005). Authors of the draft state that metadata of high quality is often not shareable and suggests sharing of metadata hinges on the following criteria: 1) proper context 2) content coherence 3) standard vocabularies 4) consistency and 5) technical conformance (p. 5). Shreeves, Riley, and Milewicz (2006) continue the discussion of what makes metadata shareable and how the focus on the local environment can limit the use of metadata for other purposes.

Shepherd (2006) details various current metadata projects funded by the United Kingdom’s Joint Information Systems Committee (JISC) focusing on interoperability. Projects listed include the ingesting of serial subscription metadata, use of XML ONIX by small publishers for licensing terms, availability of publisher/library licenses in XML format, improved access to journal table of contents directly in the OPAC, and the further development of content tagging where users mark and share resources.

How can interoperability be promoted over time? Kelly, Closier, and Hiom (2005) detail the work done on the Social Sciences, Business, and Law (SOSIG) subject gateway hosted by the University of Bristol. The authors suggest a quality assurance method for metadata to ensure that the system remains interoperable as the project matures.

Beyond the practical means of making interoperability a reality, theory is developing. To quote Howarth (2005b), “The efforts toward better theoretical modeling informed by application of standards that are, themselves, iteratively enhanced, will move the processes of electronic resource description—inherent to both bibliographic control and metadata—along the road to a potentially new role beyond the institutional boundaries of libraries, archives, and museums” (p. 51). One such attempt at a model is the MODAL Framework for Metadata Objectives, Principles, Domains, and Architectural Layout (Greenberg 2005). The model seeks to ascertain the usefulness and context of various metadata schemes.

Metadata management

As projects mature, how is metadata being managed? Westbrook (2005) suggests that library science focus on the holistic management of “activities designed to create, preserve, describe, maintain access and manipulate metadata, MARC and otherwise, that may be owned, aggregated, or distributed by the managing institution” (p. 6). In a thought provoking article, Kurth, Ruddy, and Rupp (2004) further explore what metadata management means across an entire library and begin the discussion of how to address existing MARC records, mappings, and transformations in the context of digital collections at Cornell University.

Penn State University involves a wide range of specialties in the membership of their Digital Project Managers Plus Group which is responsible for metadata projects in the libraries (Ma 2006). Steps within their project management approach include: analyzing metadata requirements, adoption of metadata schema, creation of metadata content, delivery/access, evaluation of metadata, and sustaining of metadata maintenance. What makes the process unique is the early effort and emphasis placed on the management and upkeep of the metadata over its life.

Interest in automated metadata generation remains steady. Mitchell (2006) describes iVia and Data Fountains, two open-source options that include semi-automated and fully automated metadata generation utilities, metadata extraction, a means of inserting a controlled vocabulary, rich text identification and extraction, and discovery tools. The sheer volume of digital resources and their need for descriptive and administrative metadata would greatly benefit from further automation of metadata creation and enrichment.

Administrative metadata

As standards for descriptive metadata have matured, attention has begun to shift to other types of metadata such as administrative metadata. Administrative metadata involves ownership, object history, creation date, and other related production information (Intner, Lazinger, and Weihs 2006, p. 12).

Administrative metadata is crucial to the maintenance of digital intellectual property rights and should be easily revealed. Moreover, how libraries actively manage their electronic licenses becomes even more important as licensing complexities multiply. Farb and Riggio (2004) expound the need for metadata to address electronic resource licensing and show how the Dublin Core and ONIX schemas currently fall short of the needs for licensing metadata. Farb and Riggio (2004) also briefly discuss the work of the Digital Library Federation Electronic Resource Management Initiative on development of rights metadata.

Preservation metadata

Often considered a type of administrative metadata, preservation metadata is coming into its own and has received more attention in the literature recently. Knight (2005) discusses the experiences of the National Library of New Zealand

and their early work on the development of preservation metadata. Knight raises some interesting questions, such as what role will automation play in the creation of preservation metadata, when should it be captured, who is doing the updating, and how it should be done (p. 97).

Lavoie and Gartner (2005) provide a very readable overview of preservation metadata and why it is critical for digital objects, as well as a history of the development of a shared preservation element set and data dictionary called Preservation Metadata: Implementation Strategies or PREMIS. Caplan and Guenther (2005), co-chairs of PREMIS, the OCLC/RLG Working Group on Preservation Metadata Implementation Strategies, provide greater detail on the creation of PREMIS, which comprises basic preservation metadata elements tied to an XML-data dictionary to avoid being platform specific. With digital objects and digital projects growing exponentially, it becomes even more critical to work on the total framework and the management of the metadata being created.

Areas of future research

1. Metadata management now and in the future

Metadata management implies continued maintenance over the life of a project. Kurth, Rupp, and Ruddy (2004) focus on the library-wide implications and the need to look at all metadata generated, be it MARC or one of the other numerous schemas currently being used. Metadata with regards to digital collections is often permanent. What strategies should be developed in order to manage metadata effectively now and in the future? Ma (2006) mentions a holistic approach to digital projects. More work is needed on the life cycle of metadata and how to incorporate it into the total workflow and out of the exclusive domain of special projects.

2. Interoperability, integration, and “shareable” metadata

Numerous articles have described the process of mapping metadata from one schema to another (Carini and Shepherd 2004). Beyond the practical tasks of mapping and transformation, Greenberg (2005) suggested further research is needed looking at how metadata schemas operate “in order to understand their place in the larger context of information organization, management, and access ... A framework is needed to study the full extent of and functionalities supported by metadata schemes” (p. 19). No metadata schema is applied in total isolation. Greenberg’s MODAL framework provides a means of further study of various schemas. True interoperability and sharing of metadata relies heavily on an understanding of the whole, i.e. how various metadata schemas further retrieval. While much attention has been focused on the parts, additional research is required in how the sum operates and the degree to which retrieval is facilitated.

3. Further development and use of non-descriptive metadata

Much of the literature has focused on descriptive metadata. Now that descriptive metadata has matured, attention is turning to administrative and preservation metadata. Administrative metadata attempts to document the restrictions and

intellectual rights of born digital objects. What is being done with administrative metadata and electronic resource management (ERM) systems? In order for today's digital objects to be viable over time, efforts must be made to document their use and any restrictions. The California Digital Library's CopyrightMD is one attempt at addressing rights metadata. The concept of robust administrative metadata is extremely important when considering life of a digital object.

Preservation issues need to be addressed in all digital libraries (e.g., normalization, migration on demand, format migration, etc.) (Caplan and Guenther 2005). One recent development in the arena of preservation metadata is PREMIS. Additional research is needed on how PREMIS being implemented by libraries and other cultural institutions with a special interest in how to automate the capture of preservation metadata.

Works cited

Best Practices for Shareable Metadata (draft): Part of the Best Practices for OAI Data Provider Implementations and Shareable Metadata. August 2005. Viewed June 10, 2007,

<http://oai-best.comm.nsd.org/cgi-bin/wiki.pl?PublicTOC>

California Digital Library. CopyrightMD Schema, Frequently Asked Questions. March 24, 2006. Viewed June 10, 2007,

<http://www.cdlib.org/inside/projects/rights/schema/faq.html>

Caplan, Priscilla, and Rebecca Guenther. 2005. Practical preservation: The PREMIS experience. *Library Trends*, 54 (1): 111-124.

Carini, Peter, and Kelcy Shepherd. 2004. The MARC standard and encoded archival description. *Library Hi Tech*, 22 (1): 18-27.

Chan, Lois Mai. 2005. Metadata interoperability: A study in methodology. *Chinese Librarianship, an International Electronic Journal*, 19.

Chan, Lois Mai, and Marcia Lei Zeng. 2006. Metadata interoperability and standardization: A study of methodology part I. *D-Lib Magazine*, 12:6.

Chopey, Michael A. 2005. Planning and implementing a metadata-driven digital repository. *Cataloging & Classification Quarterly*, 40 (3/4): 255-286.

Coyle, Karen. 2004. Future considerations: The functional library systems record. *Library Hi Tech*, 22 (2): 166-174.

Coyle, Karen. 2005. Understanding metadata and its purpose. *The Journal of Academic Librarianship*, 31 (2): 160-163.

Cundiff, Morgan V. 2004. An introduction to the Metadata Encoding and Transmission Standard (METS). *Library Hi Tech*, 22 (1): 52-64.

DeZelar-Tiedman, Christine. 2004. Crashing the party: Catalogers as digital librarians. *OCLC Systems & Services*, 20 (4): 145-147.

Eden, Bradford Lee. 2004. Metadata and librarianship: Will MARC survive? *Library Hi Tech*, 22 (1): 6-7.

El-Sherbini, Magda, and George Klim. 2004. Metadata and cataloging practices. *The Electronic Library*, 22 (3): 238-248.

Farb, Sharon E., and Angela Riggio. 2004. Medium or message? A new look at standards, structures, and schemata for managing electronic resources. *Library Hi Tech*, 22 (2): 144-152.

Goldsmith, Beth, and Frances Knudson. 2006. Repository librarian and the next crusade. *D-Lib Magazine*, 12:9.

Greenberg, Jane. 2005. Understanding metadata and metadata schemes. *Cataloging & Classification Quarterly*, 40 (3/4): 17-36.

Hillmann, Diane I., and Elaine L. Westbrook, eds. 2004. *Metadata in practice*. Chicago: American Library Association.

Howarth, Lynne C. 2005. Enabling metadata: Creating core records for resource discovery, 2004 update on activities of the IFLA Cataloguing Section Working Group on the Use of Metadata Schemas. *International Cataloguing and Bibliographic Control*, 34 (1): 14-17.

Howarth, Lynne C. 2005b. Metadata and bibliographic control: Soul-mates or two solitudes? *Cataloging & Classification Quarterly*, 40 (3/4): 37-56.

Howarth, Lynne C. 2004. Metadata schemas for subject gateways. *International Cataloguing and Bibliographic Control*, 33 (1): 8-12.

Intner, Sheila S., Susan S. Lazinger, and Jean Weihs. 2006. *Metadata and its impact on libraries*. Westport, Conn.: Libraries Unlimited.

Kelly, Brian, Amanda Closier, and Debra Hiom. 2005. Gateway standardization: A quality assurance framework for metadata. *Library Trends*, 53 (4): 637-650.

Knight, Steve. 2005. Preservation metadata: National Library of New Zealand experience. *Library Trends*, 54 (1): 91-110.

Kurth, Martin, David Ruddy, and Nathan Rupp. 2004. Repurposing MARC metadata: Using digital project experience to develop a metadata management design. *Library Hi Tech*, 22 (2): 153-165.

Lavoie, Brian, and Richard Gartner. 2005. Technology Watch Report: Preservation Metadata. Accessed June 10, 2007, <http://www.dpconline.org/docs/reports/dpctw05-01.pdf>.

Library of Congress. Metadata Authority Description Schema (MADS). Accessed June 10, 2007, <http://www.loc.gov/standards/mads/>.

Library of Congress. PREMIS: Preservation Metadata Maintenance Activity. Accessed June 10, 2007, <http://www.loc.gov/standards/premis/>.

- Ma, Jin. 2006. Managing metadata for digital projects. *Library Collections, Acquisitions, & Technical Services*, 30:3-17.
- McCallum, Sally H. 2004. An introduction to the Metadata Object Description Schema (MODS). *Library Hi Tech*, 22 (1): 82-88.
- Medeiros, Norm. 2006. On the Dublin Core front: Metadata in a global world. *OCLC Systems & Services*, 22 (2): 89-91.
- Mitchell, Steve. 2006. Machine-assisted metadata generation and new resource discovery: Software and services. *First Monday*, 11:8.
- Polydoratou, Panayiota. 2006. Using web logs transactions to assess a metadata registry system's use: The case for MetaForm. *OCLC Systems & Services*, 22 (1): 67-79.
- Robertson, R. John. 2005. Metadata quality: Implications for library and information science professionals. *Library Review*, 54 (5): 295-300.
- Shepherd, Peter T. 2006. International dateline: The importance of metadata and interoperability. *Against the Grain*, 18:2, 80-81.
- Shreeves, Sarah L., Jenn Riley, and Liz Milewicz. 2006. Moving towards shareable metadata. *First Monday*, 11:8.
- Smiraglia, Richard P. 2005. Introducing metadata. *Cataloging & Classification Quarterly*, 40 (3/4): 1-15.
- Tennant, Roy. 2004. A bibliographic metadata infrastructure for the twenty-first century. *Library Hi Tech*, 22 (2): 175-181.
- Wagner, Harry, and Stuart Weibel. 2005. The Dublin Core Metadata Registry: Requirements, implementation, and experience. *Journal of Digital Information*, 6 (2). Accessed June 10, 2007, <http://jodi.tamu.edu/Articles/v06/i02/Wagner/>.
- Weibel, Stuart. 2005. Border crossings: Reflections on a decade of metadata consensus building. *D-Lib Magazine*, 11 (7/8). Accessed June 10, 2007, <http://www.dlib.org/dlib/july05/weibel/07weibel.html>.
- Westbrooks, Elaine L. 2005. Remarks on metadata management. *OCLC Systems & Services*, 21 (1): 5-7.
- Zeng, Marcia Lei, and Lois Mai Chan. 2006. Metadata interoperability and standardization: A study of methodology part II. *D-Lib Magazine*, 12:6.

Prepared by:
Michelle R. Turvey-Welch
Original Cataloger
Kansas State University
mturvey@ksu.edu