

Metadata: A Primer for Indexers

BY DEBBIE OLSON

Debbie Olson provides freelance indexing, library and archival services and lives in Syracuse, NY. She is the current Webmaster for the Web Indexing SIG and has recently audited classes in information architecture and metadata from Syracuse University. She can be reached at dbolso@aol.com or www.olson-info.com.

Introduction

Metadata. It's a term being used widely and often in the library (digital libraries, cataloging), business (knowledge management), and web design (information architecture and taxonomy development) worlds. Its most basic definition – one that you are probably already familiar with – is “information about information.” It's the mantra of the information world. But here's a more detailed definition from the National Information Standards Organization (NISO) where the key to metadata is that it is “structured” information (2004):

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.¹

Its nuances now cast a broader net in terms of meaning and application. It's partially about a change in terminology stemming from a change in technology but it's also about a new way of looking at information. And as standards and best practices evolve, information can be more easily accessed and shared across communities and platforms.

Indexers are already creators of metadata and have important skills and experience to bring to the broader discussion. We can start by developing a basic understanding of the definitions, standards, schemas, applications, and best practices associated with the world of metadata.

Applications for Metadata

Metadata creation is not limited to text-based material. It can be applied to three-dimensional objects, images, audio, video materials, geospatial data, PowerPoint presentations, lesson plans, websites, computer file systems, and just about anything we want in order to find it faster and more efficiently in the future. A few examples include:

- The traditional library card catalog – certainly one of the earliest, mainstream forms of meta-

data that fits the definition.

- “Digital libraries” made up of born-digital information or materials originally in another format brought together online via scanning or other technologies. Such initiatives are seen in the traditional library environments but organizations also use digital library concepts to document their in-house information.
- The information contained in the `<head></head>` element of a web page, e.g. keywords.
- The technical details associated with photos from a digital camera, e.g. width, height, pixels per inch, resolution.
- The information about music downloaded from an online source e.g. artist, length, genre.
- The properties of a text document, e.g. type of file, location, creator.

Some types of metadata may be automatically embedded in a record such as in the case of photos taken from a digital camera or in downloaded music. It can also be proprietary, as in the case of the metadata stored in a text editor, or non-proprietary, such as that stored in the header of an HTML document. Other metadata is added by humans. Tagging photos on a gallery website such as Flickr is one example.²

Types of Metadata

Metadata is generally broken down into four or five categories. *Administrative* metadata includes rights and preservation metadata (sometimes separate) and contains information regarding the creation, management, and accessibility of the material. *Technical* metadata (sometimes included with administrative) contains information about how the item or piece of information was generated or transferred, e.g. camera settings for a photograph that are automatically embedded in the file/object. *Structural* metadata refers to how the information or object is to be compiled, e.g. filenames for chapters of a

book. *Descriptive* metadata, the type of metadata we're most familiar with and indexers are most likely to work with, includes elements such as titles, genres, authors/creators, and subjects.³

Metadata Standards and Schemas

Metadata standards and schemas provide a way to format and exchange metadata and improve *interoperability*, or the ability to be understood and shared across information communities and platforms. *Metadata standards* address how to describe content and what must be included, such as a title, creator, date, and subject. Standards also provide the rules, or *syntax*, for how *values*, e.g., the name of a particular title or a subject heading, are constructed or formatted. *Metadata schemas*, as opposed to schemas for encoding material for the Web such as XML, include specific sets of elements, e.g., title, creator, date, and subject.

In the library community, the *Anglo-American Cataloging Rules* (AACR) provides the rules to describe the content of a bibliographic record while *MAchine Readable Cataloging* (MARC) provides the schema, or set of elements, to markup and encode this bibliographic information in an electronic format, e.g., 245 is the title field. In the archival community, *Describing Archives: A Content Standard* (DACS), provides the standards for describing archival materials while *Encoded Archival Description* (EAD) provides the set of elements necessary to markup and encode the material, such as the components of a finding aid, for an online environment.

Current metadata standards and schemas were developed and became more widespread primarily in the mid-1990s with the creation of the Dublin Core Metadata Element Set (DCMES). Sometimes simply known as "Dublin Core," it is named after Dublin, Ohio, the location where the original workshop convened and where OCLC, the library cataloging services company, is headquartered. The initiative called for a set of categories that could be used to describe electronic resources and involved professionals in the fields of computer science, librarianship, archives and indexing.⁴

Objectives of the workshop included creating a core set of elements and promoting understanding among stakeholders, information communities and users.⁵ The result was a basic set of fifteen elements: contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title and type.⁶ Each element is optional and repeatable and provides enough simplicity and flexibility to be adopted by a variety of communities.

An example of DCMES in the business com-

munity can be viewed in the <head></head> section of the page info and page source codes of the home page for IBM.⁷ Dublin Core is declared as the *namespace*, where the name of the schema and the web location (URL) where it resides is given. For example, IBM uses six Dublin Core elements that are noted by the "dc" prefix: dc.publisher, dc.language, dc.date, dc.rights, dc.subject, and dc.type. W3C-DTF, a syntax standard for expressing the date and time, is used for configuring the date and dc.subject refers to internal IBM taxonomies.

DCMES, however, has become too simplistic for some communities. The fifteen elements were refined and expanded with additional properties but other standards and schemas have been developed too. A few examples include: *Categories for the Description of Works of Art* (CDWA) and the *VRA Core Categories* (VRA) for artwork and other visual media; *IEEE Learning Object Metadata* (LOM) and *Gateway to Educational Materials* (GEM) for educational materials; *Encoded Archival Description* (EAD) for archival finding aids; *Content Standards for Digital Geospatial Metadata* (CSDGM) for scientific material; and *Online Information eXchange* (ONIX) and the *Text Encoding Initiative* (TEI) for booksellers and publishers. Elements from various schemas may also be combined to create an *application profile*, a formal set of guidelines on how a particular community is using the elements of one or several schemas.

The public sector is also involved and invested in developing standards and schemas to manage the ever-expanding amount of online government information. The Australian Government Locator Service (AGLS), the e-Government Metadata Standard (e-GMS) (UK), and the Government of Canada Records Management Metadata Standard (GC RMMS) are examples. The U.S. federal government recommends the use of metadata as a part of its best practices for government websites.⁸ Many of these public standards are based on DCMES.

The Role of Controlled Vocabularies

Specific controlled vocabularies, such as the *Sauri* and *Name and Subject* authorities, may either be recommended for use with a particular standard or may need to be located, refined, or created from a variety of vocabularies or resources, especially if the terminology is localized or emerging. The Getty Institute's *Art and Architecture Thesaurus* (AAT)[®], *Union List of Artist Names* (ULAN)[®], *Thesaurus of Geographic Names* (TGN)[®], and *Cultural Objects Name Authority* (CONA)[™] are widely used as are the Library of Congress' *Name and Subject Authorities*.⁹ Other vocabularies, such as the *DCMI Type Vocabulary*, provide a list of values for a specific element in DCMES, in this case,

the genre term to input in the "type" element. It can be adapted or modified for non-DCMES projects.

The Tools of Metadata Creation

Metadata can be input by hand or with the assistance of tools such as templates and editors. Schemas may also have tools created especially for them to assist with creating and encoding metadata for the Web. One example is DC-dot, a Web-based application that will generate DCMES metadata in HTML or XML for embedding in the <head></head> of a Web page.¹⁰

There are also *crosswalks*, tools to assist with converting one schema to another through mapping one set of elements to another or via an intermediary, e.g., one schema is mapped to DCMES then mapped to another schema from DCMES. It is important to keep in mind that with any of these tools a certain amount of editing will be necessary to meet the needs of a particular project or community.

Metadata may also be created through a content management system (CMS) which provides for the authoring, publishing, archiving, and searching of documents. Many corporations are using CMSs to manage websites and internal information¹¹ and many cultural institutions are using digital management software packages, e.g., CONTENTdm[®], to catalog their collections.¹² Open source (free) systems are also available to manage content and metadata.

Indexers as Metadata Creators

In expanding our knowledge and vocabulary to include the world of metadata we can take a step towards not only maintaining, but more importantly, advancing our role in the information world in our own eyes and those of the wider information community. We can begin by familiarizing ourselves with the basics of metadata and its applications, standards, schemas, best practices, and tools available to assist us with its creation.

Through attending workshops and participating in ventures beyond the indexing world we can actively bring our indexing knowledge and expertise to the broader information table and open ourselves and our profession to new opportunities for collaboration while also expanding and developing new markets for our services.¹²

End Notes

- 1 Understanding Metadata. Bethesda, MD: National Information Standards Organization. 2004. www.niso.org/publications/press/UnderstandingMetadata.pdf
- 2 For an example of the application of metadata in regards to photo gallery sites, see *Fred Brown's* article, "Metadata Goes Mainstream," KnowGene-

- sis International Journal for Technical Communication. Vol. 2, Issue 1 (March 2007), pp 9-10. www.allegrotechindexing.com/metadata.pdf
- ³ Understanding Metadata. Bethesda, MD: National Information Standards Organization. 2004. www.niso.org/publications/press/UnderstandingMetadata.pdf
- ⁴ Dublin Core Metadata Initiative (DCMI). www.dublincore.org/workshops/dc1/. Although the indexing community is listed as involved in the workshop, ASI was not listed on the registration list (also available online at this address). Accessed December 12, 2008.
- ⁵ Dublin Core Metadata Initiative (DCMI). www.dublincore.org/workshops/dc1/general.shtml. Accessed December 12, 2008.
- ⁶ For details visit: www.dublincore.org/documents/dces/.
- ⁷ Visit IBM at www.ibm.com/us/. Accessed December 12, 2008.
- ⁸ Federal Web Content Manager's Advisory Council (US). www.usa.gov/webcontent/managing_content/organizing/metadata.shtml.
- ⁹ For more information on controlled vocabularies, their use, creation and links, visit ASI's Taxonomies and Controlled Vocabulary SIG site at www.taxonomies-sig.org/.
- ¹⁰ DC-dot. www.ukoln.ac.uk/metadata/dcdot/.
- ¹¹ For more information on content management systems (CMS), see Fred Leise's article, "Metadata and Content Management Systems: An Introduction for Indexers," *The Indexer*, Vol. 24, no. 2 (October 2004), pp. 71-74. Available online at: www.contextualanalysis.com/publications/Indexer_2004_02_Leise%20-%20Final.pdf.
- ¹² For examples of digital library collections using metadata, visit the CONTENTdm in action web page at: www.oclc.org/contentdm/collections/default.htm. You will be able to view the item and its metadata once you choose an individual item within a collection.

Metadata in the Library, Business, Information architecture or Indexing professions

- Baca, Murtha, ed. *Introduction to Metadata: Pathways to Digital Information*, 2nd edition (online version 3.0), 2008. Los Angeles, CA: The J. Paul Getty Trust. Available at: www.getty.edu/research/conducting_research/standards/intrometadata/
- Brand, Amy, et. al. *Metadata Demystified: A Guide for Publishers*. The Sheridan Press & NISO Press, July 2003. Available at: www.niso.org/publications/press/Metadata_Demystified.pdf.
- Browne, Glenda and Jon Jermy. *The Indexing Companion*. Cambridge, England: Cambridge University Press, 2007.
- Foulonneau, Muriel and Jenn Riley. *Metadata for Digital Resources: Implementation, Systems Design and Interoperability*. Oxford, England: Chandos Press, 2008. (See Heather Hedden's review in this issue of *Key Words*)
- Lambe, Patrick. *Organising Knowledge: Taxonomies, Knowledge and Organizational Effectiveness*. Oxford, England: Chandos Press, 2007.
- Morville, Peter and Louis Rosenfeld. *Information Architecture for the World Wide Web*, 3rd ed. Sebastopol, CA: O'Reilly Media, Inc., 2007.

- Taylor, Arlene G. *The Organization of Information*, 2nd edition. Westport, CT: Libraries Unlimited, 2004.
- Wyman, Pilar L. "Navigating the Future of Technical Communication: STC's 51st Annual Conference," *Key Words*, vol 12, no. 3 (July-September 2004), pp. 100-104. This conference report summarizes two sessions that discuss the role of indexers in metadata creation.
- Zeng, Marcia Lei and Jian Qin. *Metadata*. New York: Neal-Schuman Publishers, Inc. 2008.

Metadata Standards and Schemas

- Australian Government Locator Service (AGLS). www.nla.gov.au/metadata.html.
- Categories for the Description of Works of Art (CDWA). The Getty. www.getty.edu/research/conducting_research/standards/.
- Content Standard for Digital Geospatial Metadata (CSDGM). Federal Geographic Data Committee. www.fgdc.gov/metadata/csdgm/.
- e-Government Metadata Standard (e-GMS) (UK). www.govtalk.gov.uk/schemasstandards/metadata_document.asp?docnum=768.
- Federal Web Content Manager's Advisory Council (US) www.usa.gov/webcontent/managing_content/organizing/metadata.shtml.
- Gateway to Educational Materials (GEM). Sponsored by the National Education Association. www.thegateway.org/about/gemingeneral/about-gem/.
- Government of Canada Records Management Metadata Standard (GC RMMS). www.collectionscanada.gc.ca/government/products-services/007002-5001-e.html.
- Learning Object Metadata (LOM). IEEE. itsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.
- Library of Congress. www.loc.gov/standards/.
- United Kingdom. The e-Government Metadata Standard (e-GMS). www.govtalk.gov.uk/schemasstandards/metadata.asp.
- Visual Resources Association (VRA) Core Categories. www.vraweb.org/projects/vracore4/index.html.

Encoding Standards

- Encoded Archival Description (EAD). Library of Congress. www.loc.gov/ead/.
- Dublin Core Metadata Initiative (DCMI). dublincore.org/.
- MACHine Readable Cataloging (MARC). www.loc.gov/marc/.
- ONline Information eXchange (ONIX). Maintained by EDItEUR with the Book Industry Communication (UK) and the Book Industry Study Group (US). www.editeur.org/onix.html.
- Text Encoding Initiative (TEI). TEI Consortium, eds. www.tei-c.org/index.xml.

Content Standards

- Anglo-American Cataloging Rules (AACR). Chicago: American Library Association, 2002.
- Describing Archives: A Content Standard (DACS). Chicago: Society of American Archivists, 2004. www.archivists.org/catalog/pubDetail.asp?objectID=1279.
- Digital Library for Earth Systems Education. www.dlese.org/library/index.jsp.

- Visual Resources Association Foundation. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images (CCO)*. www.vrafoundation.org/ccoweb/index.htm.

Syntax Standards

- Date and Time Formats (W3C-DTF). www.w3.org/TR/NOTE-datetime.
- ISO Country Code List. ISO 3166-1193 (E). www.iso.org/iso/country_codes/iso_3166_code_lists/english_country_names_and_code_elements.htm.

Controlled Vocabularies

- DCMI Type Vocabulary. dublincore.org/documents/dcmi-type-vocabulary/.
- Getty Institute Vocabularies. www.getty.edu/research/conducting_research/vocabularies/.
- Library of Congress Name and Subject Authorities. authorities.loc.gov/.
- National Library of Medicine. Medical Subject Headings (MeSH). www.nlm.nih.gov/mesh/.

Metadata Tools

- DC-dot. www.ukoln.ac.uk/metadata/dcdot/.
- Dublin Core tools and software. dublincore.org/tools/.
- oxygen (XML Editor). SyncRO soft Ltd. www.oxygenxml.com/xml_schema_editor.html

Content Management Systems (CMSs)

- CONTENTdm from OCLC (digital collections management). www.contentdm.com/.
- Drupal (free). www.drupal.org/.
- Joomla (free). www.joomla.org/about-joomla.html.