Metadata 101

NETSL Spring Conference, April 15, 2010

Kelcy Shepherd, Digital Interfaces Librarian, UMass Amherst
kshepher@library.umass.edu

Introduction to Metadata

Metadata

Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.

-NISO, Understanding Metadata

▶

Basic Metadata Concepts & Principles

Metadata is…

▶ a digital object

Metadata can…

▶ be applied at varying levels of aggregation

▶ be embedded within a digital resource or stored separately

Metadata *should*…

▶ conform to community-based standards

▶ be appropriate to the materials being described and the needs of users

▶ support interoperability

▶

Why Do We Need Metadata?

▸ Facilitates and expands access

▸ Enables evaluation and selection

▸ Supports navigation of digital objects

▸ Assists with management of digital objects

▸

Metadata Components

Metadata Components

▸ Elements and structure
▸ Content
  ▸ Values
▸ Format

Elements and Structure

▸ Data elements (fields) and their relationships
▸ Examples:
  ▸ Machine Readable Cataloging (MARC)
  ▸ Metadata Object Description Schema (MODS)
  ▸ Visual Resources Association (VRA)Core
  ▸ Categories for Descriptions of Works of Art (CDWA)
  ▸ Encoded Archival Description (EAD)

Content Standards

▸ Rules for the order and form of the data itself
▸ Examples:
  ▸ Anglo-American Cataloguing Rules (AACR2)
  ▸ Resource Description and Access (RDA)
  ▸ Cataloging Cultural Objects (CCO)
  ▸ Describing Archives: A Content Standard (DACS)

▸

Data Value Standards
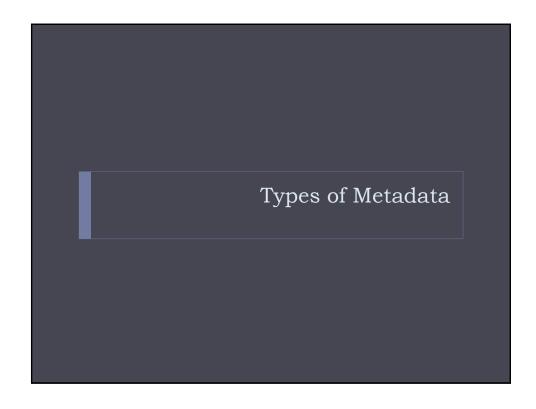
▸ Controlled list of values for specific fields
▸ Examples:
  ▸ Library of Congress Subject Headings (LCSH)
  ▸ Library of Congress Name Authorities (LCNA)
  ▸ Union List of Artists' Names (ULAN)
  ▸ Art and Architecture Thesaurus (AAT)
  ▸ Thesaurus of Geographic Names (TGN)

▸

Format

▸ Specifies data encoding
▸ Examples:
  ▸ ISO 2709
  ▸ eXtensible Markup Language (XML)

▸

Putting It All Together

Resource	Elements & Structure	Content	Format
Book	MARC	AACR2	ISO 2709, XML
Digital Object	MODS	AACR2	XML
Archival Collection	MARC	DACS	ISO 2709, XML
Archival Collection	EAD	DACS	XML
Image	VRA Core 4.0	CCO	XML
Painting	CDWA/CDWA Lite	CCO	XML
Digital Object	DC	Local Guidelines	(not specified, could be XHTML, XML, RDF)

▸

Types of Metadata

Types of Metadata

▸ Descriptive
▸ Structural
▸ Use
▸ Administrative
  ▸ Technical
  ▸ Preservation
  ▸ Rights

▸

Descriptive Metadata

- Information about the intellectual content and physical format
- Supports discovery, identification, and selection
- Standards:
  - Dublin Core
  - MARC
  - MODS
  - VRA Core
  - CDWA
  - EAD

Structural Metadata

- Information about the relationships among individual components of a complex digital object
- Supports navigation
- Includes data related to:
  - Individual files
  - Physical and/or intellectual structure
  - Behaviors
- Standards:
  - METS
    http://www.loc.gov/standards/mets/

Use Metadata

▸ Information about how and how much the object has been used

▸ Supports assessment

▸ Includes data related to:

  ▸ Circulation

  ▸ Exhibits

  ▸ User tracking and behavior

▸

Technical Metadata

▸ Information about the technical processes used to produce or required to use a digital object

▸ Supports quality assessment, accurate rendering, preservation

▸ Includes data related to:

  ▸ Hardware and software

  ▸ Digitization

  ▸ Authentication and security details

▸ Standards:

  ▸ Metadata for Images in XML (MIX)
    http://www.loc.gov/standards/mix/

  ▸ Technical Metadata for Text (textMD)
    http://www.loc.gov/standards/textMD/

▸

Preservation Metadata

- Information about preservation management of digital materials
- Supports long-term retention and accessibility
- Includes data related to:
  - The object's creation
  - Changes to the digital object
  - Chain of custody
  - Technical requirements for access
- Standard:
  - PREservation Metadata: Implementation Strategies (PREMIS)
    http://www.loc.gov/standards/premis/

-

Rights Metadata

- Information about rights related to access and use
- Supports appropriate use by end users
- Includes data related to:
  - Copyright
  - Licensing
  - Terms and conditions of use
- Standards:
  - CopyrightMD
    http://www.cdlib.org/inside/projects/rights/schema/
  - METSRights
    http://www.loc.gov/standards/mets/news080503.html

-

Descriptive Metadata:
Dublin Core, MODS, VRA Core, EAD

Dublin Core

▸ Designed to support description of a wide range of resources
▸ Two levels: Simple and Qualified
▸ No format specified
▸ Documentation:
  ▸ http://dublincore.org/
▸ Projects:
  ▸ Maine Memory Network
    http://www.mainememory.net/
  ▸ Connecticut History Online
    http://www.cthistoryonline.org/cdm-cho/index.html
▸

Dublin Core Record

Title	Subway Caves
Creator	Eastman Studio
Link	http://www.cripple-creek.org/cgi-bin/photograph.cgi?id=2000113
Publisher	Eastman Studio
Description	Postcard Front shows cave tunnel. White lettering at bottom reads "Subway Caves, Lassen Volcanic National Park, Calif. c 1939 - Eastman Studio B-848" Back has black outlined letters reading: "photo POST CARD". Pencil handwriting reads: "June 23-50 Just leaving here couldnt catch any fish, so Bill says he wans to go home hope you are O.K. by Now, lots of luck Bill & Evie Mc" Address in pencil reads "Mrs. G. Harris Box 123 Pismo Beach California" Postal stamp reads "Hat Creek Calif. June 24 1950 A.M." with green one cent stamp.
Date Original	1948? - 1952?
Date Digital	2000-12-06
Subject(s)	California Harris, Mrs.. G. Stumpff, Mrs.. Sam
Type	image text
Source	Glossy black and white photo on postcard photographic paper; 3 1/2" x 5 1/2"
Relation	Heritage Colorado
Format Create	image/tiff 124,351 bytes 515h x 811w pixels compression=0 quality=24bit color,300 ppi; operatingsystem=Windows 2000; Gateway E4600, Hewlett Packard ScanJet; Scanned photographs using Adobe PhotoShop, color profile
Rights	Copyright restrictions applying to use of reproduction of this image available from the Cripple Creek District Museum.
Project	Cripple Creek District Museum and Franklin Ferguson Memorial Library

Source: Colorado Digitization Program

MODS

‣ A subset of MARC elements, but with language-based tags
‣ Designed to be particularly applicable to digital resources
‣ XML format
‣ Documentation:
  ‣ http://www.loc.gov/standards/mods/
‣ Projects:
  ‣ Library of Congress, I Hear America Singing
    http://www.loc.gov/rr/perform/ihas/ihashome.html
  ‣ Calisphere
    http://www.calisphere.universityofcalifornia.edu/

Visual Resources Association (VRA) Core

- Description of visual materials
- Can be used to describe works, images, and collections
- VRA Core 4.0 introduced XML format
- Documentation
  - http://www.vraweb.org/projects/vracore4/
- Examples:
  - Display: http://gort.ucsd.edu/escowles/vracore4/examples/02-display.html
  - Record: http://gort.ucsd.edu/escowles/vracore4/examples/02.xml

‣

Encoded Archival Description (EAD)

- Standard for describing archival materials
- Supports complex hierarchical structure
- XML format
- Documentation:
  - http://www.loc.gov/ead/index.html
- Projects:
  - Northwest Digital Archive http://nwda.wsulibs.wsu.edu/index.shtml
  - Online Archive of California http://www.oac.cdlib.org/

‣

And many more…

- Categories for Descriptions of Works of Art (CDWA)
  http://www.getty.edu/research/conducting_research/standards/cdwa/
- Public Broadcasting Core (PBCore)
  http://www.pbcore.org/
- Darwin Core
  http://www.tdwg.org/activities/darwincore/
- Federal Geographic Data Committee (FGDC)
  http://www.fgdc.gov/metadata/geospatial-metadata-standards
- Data Documentation Initiative(DDI)
  http://www.icpsr.umich.edu/DDI/
- ONline Information eXchange (ONIX)
  http://www.editeur.org/8/ONIX/
-

Structural Metadata:
METS

Metadata Encoding and Transcription Schema (METS)

▸ Designed for management, exchange, and display of digital objects

▸ Capacity for descriptive, structural, and administrative metadata, incorporating other metadata standards

▸ XML format

▸ Documentation:

   ▸ http://www.loc.gov/standards/mets/

▸ Projects:

   ▸ Brown University Library Center for Digital Initiatives http://dl.lib.brown.edu/

   ▸ University of Florida Digital Collections http://www.uflib.ufl.edu/ufdc/

   ▸ DRAM (Database of Recorded American Music) http://www.dramonline.org/
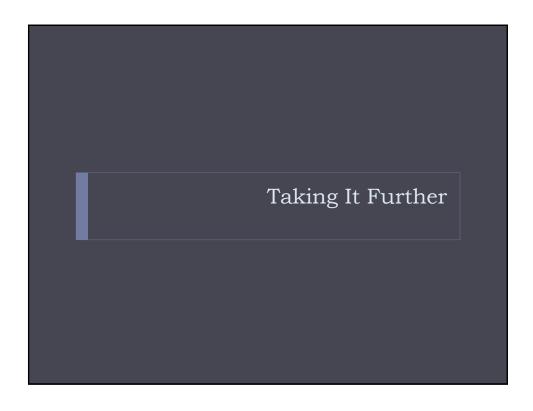
▸

Why So Many Standards?

▸ Functions

▸ Purposes

▸ Formats

▸ Audiences

▸ Communities

▸

Title Element Comparison

Standard	Excerpts from Title Definition/Rule
AACR2	•Transcribe the title proper exactly as to wording, order, and spelling, but not necessarily as to punctuation and capitalization. •If the title proper is not taken from the chief source of information, give the source of the title in a note.
CCO	•The Title element records the titles, identifying phrases, or names given to a work of art or architecture. •One of the differences between a book or article title and a title for many works of art is that printed books and journal articles generally have an inscribed title as part of the thing itself. •Titles may come from various sources. •Construct titles when necessary.
DACS	•This element provides a word or phrase by which the material being described is known or can be identified. •A title may be supplied or formal. •Supplied titles generally have two parts: the name of the creator(s) or collector(s); the nature of the materials being described.

▶

Title Element Comparison

Standard	Excerpts from Title Definition/Rule
Dublin Core	•A name given to the resource. •Typically, a Title will be a name by which the resource is formally known.
CDP Dublin Core Best Practices	•Enter multiple titles in the order in which they appear on the resource or in order of their importance. Use separate Title elements to enter multiple titles or clearly separate each entry by a semicolon and a space within an element. •Transcribe the title, if there is one, from the resource itself, such as a caption from a photograph or a title on a map. •When no title is found on the resource itself, use a title assigned by the contributing institution or found in reference sources. For more guidance in constructing titles, consult established cataloging rules such as … •Make the title as descriptive as possible, avoiding simple generic titles such as "Papers" or "Annual report."

▶

Taking It Further

Implementation

- Selecting appropriate standards
- Documentation
- Tools
- Resources

Advanced Topics
(or, some things we didn't have time for today)

- Interoperability:
    - Crosswalks and mapping
    - Shareable metadata
    - Metadata harvesting, OAI-PMH
    - Linked data
- eXtensible Markup Language (XML)
- User-contributed metadata
- Automatically-generated metadata

Sources & Additional Resources

- Baca, Murtha (editor). *Introduction to Metadata.* (2008) http://www.getty.edu/research/conducting_research/standards/intrometadata/
- Elings, Mary W. and Günter Waibel, "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives, and Museums," *First Monday* 12, no. 3 (2007). http://firstmonday.org/article/view/1628/1543
- National Information Standards Organization (NISO). (2004) *Understanding Metadata.* http://www.niso.org/standards/resources/UnderstandingMetadata.pdf
- National Information Standards Organization (NISO). (2007) *A Framework of Guidance for Building Good Digital Collections.* http://framework.niso.org/