Cataloging 101: An Introduction to Metadata

Tina Baich
ILF Annual Conference
November 17, 2010
Agenda

• Metadata in context

• Explore metadata schema

• Types of metadata

• Application Profiles
METADATA IN CONTEXT
Traditional v. Digital Libraries

Physical v. Virtual
What is a digital library?

“Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities.”

*definition from the Digital Library Federation*
What is metadata?

• data about data (or data about resources)
• “structured information that describes, explains, locates, and otherwise makes it easier to retrieve and use an information resource”*
• “value added information that is created to arrange, describe, track and otherwise enhance access to information objects”**
• What we already do!

Characteristics of metadata

- No dominant schema
- Local documentation
- Culture of sharing still developing
- Flexible
Traditional and Digital

• It’s all about access!
• Pressure to create fast and cheap description that is also rich and reusable
• Wide variety of materials and resources to describe
• Maintenance takes a back seat
## Why is metadata important?

<table>
<thead>
<tr>
<th>Discover Resources</th>
<th>Manage Documents</th>
<th>Control IP Rights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify Versions</td>
<td>Certify Authenticity</td>
<td>Indicate Status</td>
</tr>
<tr>
<td>Mark Content Structure</td>
<td>Situate Geospatially</td>
<td>Describe Processes</td>
</tr>
</tbody>
</table>
Uses of metadata

• By information specialists
  • Describing non-traditional materials
  • Cataloging websites
  • Navigating and managing digital objects
• By everyone
  • Preparing websites for search engines
  • Managing citation lists
  • Tagging (flickr, delicious, etc.)
• Social cataloging (LibraryThing, Goodreads, etc.)
METADATA SCHEMA
Metadata Schema

- Dublin Core: many fields
- MARC21: libraries
- MODS: libraries
- CSDGM: data sets, geospatial info
- GEM: educational materials
- EAD: archives
- ONIX: publishers
- PBCore: public broadcasting
- TEI: literature
- VRA Core: art, visual works
Metadata schema may include

- **Structure Standards**: an element set and their definitions
- **Syntax Standards**: the rules for tagging or encoding
  e.g. The Dublin Core schema uses XML tagging.
- **Value Standards**: acceptable values
  e.g. “Use LCSH as a controlled vocabulary for the Subject field.”
- **Content Standards**: the rules for choosing or constructing values
  e.g. “Take title from title page.”
  e.g. “Capitalize first word in title field and end with a period.”
Public buildings
Dublin Core

• Simple to use
• All elements are repeatable
• No order of elements is prescribed
• Can be mapped to other metadata standards (e.g. MARC21)
## Dublin Core Elements: Simple

<table>
<thead>
<tr>
<th>Title</th>
<th>Creator</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Contributor</td>
<td>Coverage</td>
</tr>
<tr>
<td>Subject</td>
<td>Publisher</td>
<td>Identifier</td>
</tr>
<tr>
<td>Relation</td>
<td>Rights</td>
<td>Format</td>
</tr>
<tr>
<td>Source</td>
<td>Language</td>
<td>Type</td>
</tr>
</tbody>
</table>
## Dublin Core Elements: Qualified

<table>
<thead>
<tr>
<th>Dublin Core Elements</th>
<th>Qualified Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Creator</td>
</tr>
<tr>
<td>Title.Alternative</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Date</td>
</tr>
<tr>
<td>Date.Created</td>
<td>Date.Available</td>
</tr>
<tr>
<td>Description</td>
<td>Contributor</td>
</tr>
<tr>
<td>Description.Abstract</td>
<td></td>
</tr>
<tr>
<td>Description.TableOfContents</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td></td>
</tr>
<tr>
<td>Coverage.Spatial</td>
<td></td>
</tr>
<tr>
<td>Coverage.Temporal</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Publisher</td>
</tr>
<tr>
<td>Publisher</td>
<td>Identifier</td>
</tr>
<tr>
<td>Relation</td>
<td>Rights</td>
</tr>
<tr>
<td>Relation.IsVersionOf</td>
<td></td>
</tr>
<tr>
<td>Relation.Replaces</td>
<td></td>
</tr>
<tr>
<td>Rights</td>
<td>Format</td>
</tr>
<tr>
<td>Format</td>
<td></td>
</tr>
<tr>
<td>Format.Extent</td>
<td></td>
</tr>
<tr>
<td>Format.Medium</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Language</td>
</tr>
<tr>
<td>Source.IsPartOf</td>
<td></td>
</tr>
<tr>
<td>Source.HasPart</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>Type</td>
</tr>
<tr>
<td>Type</td>
<td></td>
</tr>
</tbody>
</table>
**To metadata or not to metadata**

**Author**
Reamy, Tom

**Source**
EContent; 27 (10) Oct 2004, pp.34-38

**ISSN**
1525-2531

---

**Abstract**
Whether to add metadata to unstructured content and how much effort is really justified to do so have been raised with increasing frequency in the world of digital information. With the Dublin Core Metadata Initiative Workshop, while some participants argued for a drastic reduction in metadata efforts or at least rethinking of metadata and how to generate value. What has become increasingly clear is that metadata is not going away and there is no one silver bullet of the basic issues around adding metadata to unstructured content and explores a range of approaches that various groups and solutions are providing. Adding keywords to documents, is leading to a more sophisticated, multi-dimensional or infrastructure based approach to metadata. Illustrations are provided in the article, and a list of companies featured. (Quotes from original text)

**Features**
il.

**Language**
English

**Publication Year**
2004

**Publication Type**
Journal Article

**Shelfmark**
3659.530425

**Update**
20050524

**Accession Number**
321183
3 Viennese arias: for soprano, obbligato clarinet in B flat, and piano / G.B. Bononcini and Emperor Joseph I; edited by Colin Lawson.


1 score (12 p.) + 2 parts; 31 cm.
TYPES OF METADATA
Types of metadata

- Descriptive
- Structural
- Administrative
  - Preservation (Technical)
  - Rights Management
What is descriptive metadata?

- Information describing a resource for purposes of discovery and identification
<table>
<thead>
<tr>
<th>Element</th>
<th>Value</th>
<th>Controlled Vocab?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>52nd Street, New York, N.Y., ca. 1948</td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Gottlieb, William P.</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>1948</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>New York (N.Y.)</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Nightclubs – New York (State) – New York</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Image</td>
<td></td>
</tr>
</tbody>
</table>
What is structural metadata?

- Information indicating how to display and/or navigate a digital object (e.g. the order of pages in a chapter, order of chapters in a book)
- Users of digital collections are not aware of this type of metadata. It’s all behind the scenes.
What is administrative metadata?

- Information that helps manage a digital object
  - Preservation metadata
    - Technical characteristics (e.g. file type, file size)
    - Information about actions taken (e.g. scanner type)
  - Rights management metadata
    - Copyright, usage rights statement
[52nd Street, New York, N.Y., ca. 1948] / William P. Gottlieb [photograph]
<table>
<thead>
<tr>
<th>Element</th>
<th>Value</th>
<th>Controlled Vocab?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>52nd Street, New York, N.Y., ca. 1948</td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Gottlieb, William P.</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>1948</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>New York (N.Y.)</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Nightclubs – New York (State) – New York</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Image</td>
<td></td>
</tr>
<tr>
<td>Identifier</td>
<td><a href="http://lcweb2.loc.gov/diglib/ihas/loc.natlib.gottlieb.02771/default.html">http://lcweb2.loc.gov/diglib/ihas/loc.natlib.gottlieb.02771/default.html</a></td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>Scanner: Konica Minolta PS7000C MKII; Master image: 600 dpi TIFF; Access image: 600 dpi JPEG</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>2003-02-15</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Library of Congress Performing Arts Encyclopedia</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Library of Congress Prints &amp; Photographs Division</td>
<td></td>
</tr>
</tbody>
</table>
Controlled Vocabularies

- LCSH - general, can combine terms to make more complex headings
- TGM I and II - graphic materials, photos, prints
- AAT - architecture, paintings, sculpture, art
- TGN - place, geographic names
- DCMI Type - acceptable object types for DC Type element
- Locally-created

Resources:
- http://www.controlledvocabulary.com/
<table>
<thead>
<tr>
<th><strong>Element</strong></th>
<th><strong>Value</strong></th>
<th><strong>Controlled Vocab?</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>52&lt;sup&gt;nd&lt;/sup&gt; Street, New York, N.Y., ca. 1948</td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Gottlieb, William P.</td>
<td>Yes: LCNAF</td>
</tr>
<tr>
<td>Date</td>
<td>1948</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>New York (N.Y.)</td>
<td>Yes: LCSH</td>
</tr>
<tr>
<td>Subject</td>
<td>Nightclubs – New York (State) – New York</td>
<td>Yes: LCSH</td>
</tr>
<tr>
<td>Type</td>
<td>Image</td>
<td>Yes: DCMI</td>
</tr>
<tr>
<td>Identifier</td>
<td><a href="http://lcweb2.loc.gov/diglib/ihas/loc.natlib.gottlieb.02771/default.html">http://lcweb2.loc.gov/diglib/ihas/loc.natlib.gottlieb.02771/default.html</a></td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>Scanner: Konica Minolta PS7000C MKII ; Master image: 600 dpi TIFF ; Access image: 600 dpi JPEG</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>2003-02-15</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Library of Congress Performing Arts Encyclopedia</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Library of Congress Prints &amp; Photographs Division</td>
<td></td>
</tr>
</tbody>
</table>
METADATA PLANNING & STANDARDIZATION
Considerations when choosing a metadata schema

- What type of material is being digitized?
- How rich does the metadata need to be?
- Have the objects been previously described/cataloged?
- What is the purpose of the project?
- Who is the audience for the collection?
- Does this collection need to interact with/complement an existing collection?
Assessing materials for ease of metadata creation

- Number of items?
- Homogeneity of items?
- Foreign language?
- Published or unpublished?
- Specialist needed?
- How much information is known?
- Any existing metadata?
Assessing currently existing metadata

- Machine-readable?
- Divided into fields?
- What format?
- What content standards?
- Complete?
Assessing software capabilities

- Are there templates for standard metadata schemas?
- Can you add/remove fields to a template?
- Can you create new templates?
- Can you add additional clarifying information without creating a separate field?
  - Personal v. corporate names
  - Subject vocabulary used
- Is there an XML export? Does it produce valid records?
Why is standardization important?

• Facilitates interoperability/information exchange

• Flexibility of metadata standards require local standardization

• Allows us to create a common look for all collections OR adapt standards to an unusual collection

• Can be used to combine features of multiple metadata standards into one local standard
Application profiles

- Documents decisions made in the planning process
- Describes the set of metadata elements, policies, and guidelines defined for a particular application, implementation, or object type
- Declares the metadata terms an organization uses in its metadata
- Documents metadata standards used including schemas, controlled vocabularies, required elements, etc.
Elements Used in IUPUI University Library Digital Collections in CONTENTdm
August 2009

Based on:
- CDP Dublin Core Metadata Best Practices Version 2.1.1
- Indiana Digital Library, Metadata Best Practices For Use of Qualified Dublin Core
Both documents can be found here: \win-ub-rlbs\shared\BAMS\Metadata\Instructional_Materials

To be applied to:
Any collection added by IUPUI staff/faculty (or in collaboration with IUPUI staff/faculty) to the IUPUI Digital Collections in CONTENTdm.

**Summary Chart:** Lists elements used in IUPUI University Library Digital Collections, the corresponding Dublin Core Element, a brief definition and whether or not the element is required in all collections.

<table>
<thead>
<tr>
<th>UL Element Name</th>
<th>Maps to DC Element</th>
<th>Definition</th>
<th>Required?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Title</td>
<td>Name of the resource.</td>
<td>Yes</td>
</tr>
<tr>
<td>Item ID/Object ID</td>
<td>Identifier</td>
<td>Unique reference to the resource.</td>
<td>Yes</td>
</tr>
<tr>
<td>Description</td>
<td>Description</td>
<td>Brief account of the content of the resource.</td>
<td>Yes</td>
</tr>
<tr>
<td>Subject</td>
<td>Subject</td>
<td>Topic of the resource.</td>
<td>Yes</td>
</tr>
<tr>
<td>Geographic Location</td>
<td>Coverage.Spatial</td>
<td>Location depicted in the resource.</td>
<td>No</td>
</tr>
<tr>
<td>Author, Photographer, etc.</td>
<td>Creator</td>
<td>Entity primarily responsible for creation of the resource.</td>
<td>Yes, if available</td>
</tr>
<tr>
<td>Date</td>
<td>Date.Created</td>
<td>Date of the creation of the resource.</td>
<td>Yes, if available</td>
</tr>
<tr>
<td>Type</td>
<td>Type</td>
<td>Nature or genre of the content of the resource.</td>
<td>Yes</td>
</tr>
<tr>
<td>Owning Institution</td>
<td>Source</td>
<td>Institution that owns the physical resource.</td>
<td>Yes</td>
</tr>
<tr>
<td>Source Collection</td>
<td>Source</td>
<td>Larger physical collection of which resource is a part.</td>
<td>Yes, if applicable</td>
</tr>
<tr>
<td>Usage Rights</td>
<td>Rights</td>
<td>Information about the rights held in and over the resource.</td>
<td>Yes</td>
</tr>
<tr>
<td>Have Questions?</td>
<td>None</td>
<td>Contact information for questions about the resource.</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Sharing metadata

• Harvesting
  • Collects metadata, processes it, and stores it locally to respond to user queries
  • Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

• Federated Searching
  • Transmits user queries to multiple destinations in real time
  • Z39.50, SRU
Why share your metadata?

• Benefits users
  • Increases accessibility
  • Facilitates one-stop searching

• Benefits the institution
  • Increases exposure
  • Increases use
Before sharing

Check your metadata!

• Accurate?
• Consistent?
• Context provided?
• Does the aggregator have what it needs?

Can a stranger tell you what the record describes?
Where your metadata can go

Google Image Search

flickr

ARTstor

Open WorldCat

Yahoo!

CALIF... find the pearls

DLF Aquifer

IUPUI INDIANA UNIVERSITY–PURDUE UNIVERSITY INDIANAPOLIS

Photograph from Indiana University Charles W. Cushman Collection

Collection Registries
Resources


QUESTIONS?

Tina Baich

cbaich@iupui.edu

https://scholarworks.iupui.edu/