Introduction to Metadata

Jenn Riley
Metadata Librarian
IU Digital Library Program
Some definitions of metadata

- Data about data
- “Structured information about an information resource of any media type or format.” (Caplan)
- “Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.” (NISO)
- “A relationship that someone claims to exist between two entities” (<indecs>)
Refining a definition

- Other characteristics
  - Structure
  - Control
- Origin
  - Machine-generated
  - Human-generated
- In practice, the term often covers data and meta-metadata
Metadata and cataloging

- Depends on what you mean by:
  - metadata, and
  - cataloging!

- But, in general:
  - Metadata is broader in scope than cataloging
  - Much metadata creation takes place outside of libraries
  - Good metadata practitioners use fundamental cataloging principles in non-MARC environments
  - Metadata created for many different types of materials
What metadata is *not*

- Just a new word for cataloging
- Only for Internet resources
- Necessarily in electronic form
- Only created by professionals
- A fundamentally new idea
- A reason to forget everything we know about describing and managing resources
Some forms of metadata in cultural heritage institutions

- Catalog cards
- OPACs
- Finding aids
- Inventories
- Registers
- Slide labels
Uses of metadata in cultural heritage institutions

- Searching
- Browsing
- Display for users
- Interoperability
- Management of digital objects
- Preservation
- Navigation
Some types of metadata

<table>
<thead>
<tr>
<th>Type</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive metadata</td>
<td>Searching, Browsing, Display, Interoperability</td>
</tr>
<tr>
<td>Technical metadata</td>
<td>Interoperability, Digital object management, Preservation</td>
</tr>
<tr>
<td>Preservation metadata</td>
<td>Interoperability, Preservation</td>
</tr>
<tr>
<td>Rights metadata</td>
<td>Interoperability, Digital object management</td>
</tr>
<tr>
<td>Structural metadata</td>
<td>Navigation</td>
</tr>
</tbody>
</table>
How metadata is used
A taxonomy of standards

Some data structure standards

- For general use
  - MARC and MARCXML
  - MODS
  - Dublin Core

- More specialized for this community
  - VRA Core
  - CDWA Lite
Brief display of LC MARC record with field labels

LC Control Number: 92005291
Type of Material: Book (Print, Microform, Electronic, etc.)
Brief Description: Sandburg, Carl, 1878-1967.
- Arithmetic / Carl Sandburg : illustrated as an anamorphic adventure by Ted Rand.
- 1st ed.
- 1 v. (unpaged) : ill. (some col.) : 26 cm.
- ISBN: 0152038655 :

CALL NUMBER: PS3537.A618 A88 1993
- Copy 1
- Request in: Jefferson or Adams Bldg General or Area Studies Reading Rms
- Status: Not Charged

CALL NUMBER: PS3537.A618 A88 1993 FT MEADE
- Copy 2
- Request in: Main or Science/Business Reading Rms - STORED OFFSITE
- Status: Not Charged
“Classic” LC MARC record with field tags

01142cam 2200301 a 4500
000
001 92005291.
003 DLC
005 19930521155141.9
008 920219s1993 cauj 000 0 eng
010 |a 92005291
020 |a0152038655 |c$15.95
040 |aDLC|cDLC|dDLC
042 |alcc
050 00|aPS3537.A618|bA88 1993
082 00|a811.52|220
100 1 |aSandburg, Carl |d1878-1967.
245 10 |aArithmetic |cCarl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.
250 |a1st ed.
300 |a1 v. (unpaged) |biill. (some col.) |c26 cm.
500 |aOne Mylar sheet included in pocket.
520 |aA poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.
650 0 |aArithmetic |xJuvenile poetry.
MARCXML record

```xml
<datafield>
  <tag>245</tag>  
  <ind1>1</ind1>  
  <ind2>0</ind2>  
  <subfield code="a">Arithmetic</subfield>  
  <subfield code="c">Carl Sandburg; illustrated as an anamorphic adventure by Ted Rand</subfield>  
</datafield>  
<datafield>
  <tag>250</tag>  
  <ind1>1</ind1>  
  <ind2>0</ind2>  
  <subfield code="a">1st ed.</subfield>  
</datafield>  
<datafield>
  <tag>260</tag>  
  <ind1>1</ind1>  
  <ind2>0</ind2>  
  <subfield code="a">San Diego</subfield>  
  <subfield code="b">Harcourt Brace Jovanovich</subfield>  
  <subfield code="c">c1993</subfield>  
</datafield>  
<datafield>
  <tag>300</tag>  
  <ind1>1</ind1>  
  <ind2>0</ind2>  
  <subfield code="a">1 v. (unpaged)</subfield>  
  <subfield code="b">ill. (some col.)</subfield>  
  <subfield code="c">26 cm</subfield>  
</datafield>  
<datafield>
  <tag>500</tag>  
  <ind1>1</ind1>  
  <ind2>0</ind2>  
  <subfield code="a">One Mylar sheet included in pocket.</subfield>  
</datafield>  
<datafield>
  <tag>520</tag>  
  <ind1>1</ind1>  
  <ind2>0</ind2>  
  <subfield code="a">A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.</subfield>  
</datafield>
```

http://www.loc.gov/standards/marcxml

Slide courtesy of Jan Eklund, UC Berkeley
Some data content standards

- Anglo-American Cataloging Rules, 2nd edition (AACR2)
- Describing Archives: A Content Standard (DACS)
  - Replaces APPM
- Cataloging Cultural Objects (CCO)
Also standards for individual fields

- Controlled vocabularies
  - Names: LCNAF, ULAN, etc.
  - “Subjects”: LCSH, AAT, etc.
  - Geographic places: LCSH, TGN, etc.
- Classification schemes, e.g., LCC
- Encoding schemes, e.g., W3CDTF

Some fields may not be covered by a content standard, or you may choose to override the guidance of a content standard for a particular field.
Creating descriptive metadata

- “Digital library” systems
  - ContentDM
  - ExLibris Digitool
  - Greenstone
- Library catalogs
- Spreadsheets & databases
- XML
Challenges for descriptive VR metadata

- No infrastructure for copy cataloging
- Often describe unique rather than mass-produced works
- Description tends to be more subjective than traditional library cataloging
- Many whole/part and derivative relationships
- Generally describe surrogates rather than actual works
Work vs. image

Work = image content

Agent = Auguste Rodin, sculptor
Title = The Thinker
Work type = sculpture
Descriptive metadata

Museum inventory # = R9876
Exhibition status = on display
Administrative metadata

Image = view of a work

Agent = Joe Dokes, photographer
Title = Full view from front left
Work type = color transparency
Descriptive metadata

Call number = 275.R692
Class = French Modern Sculpture
Accession # = 654321
Administrative metadata

Measurements = 350 kb
Work type = digital image
Technical metadata

6/25/07
SEI

Slide courtesy of Jan Eklund, UC Berkeley
The role of community

- There’s no one “right way” to record descriptive metadata
- But we shouldn’t each have to start from scratch with our own planning
- Collaboration within defined communities is the key to success