Alphabet Soup: Choosing Among DC, QDC, MARC, MARCXML, and MODS

Jenn Riley
IU Metadata Librarian
DLP Brown Bag Series
February 25, 2005
Descriptive metadata

- Enables users to find relevant materials
- Used by many different knowledge domains
- Many potential representations
- Controlled by
  - Data structure standards
  - Data content standards
  - Data value standards
Some data structure standards

- Dublin Core (DC)
  - Unqualified (simple)
  - Qualified
- MAchine Readable Cataloging (MARC)
- MARC in XML (MARCXML)
- Metadata Object Description Schema (MODS)
How do I pick one?

- Genre of materials being described
- Format of materials being described
- Nature of holding institution
- Robustness needed for the given materials and users
- What others in the community are doing
- Describing original vs. digitized item
- Mechanisms for providing relationships between records
- Plan for interoperability, including repeatability of elements
- Formats supported by your delivery software
- More information on handout
Dublin Core (DC)

- 15-element set
- National and international standard
  - 2001: Released as ANSI/NISO Z39.85
  - 2003: Released as ISO 15836
- Maintained by the Dublin Core Metadata Initiative (DCMI)
- Other players
  - DCMI Working Groups
  - DC Usage Board
DCMI mission

- The mission of DCMI is to make it easier to find resources using the Internet through the following activities:
  - Developing metadata standards for discovery across domains,
  - Defining frameworks for the interoperation of metadata sets, and,
  - Facilitating the development of community- or disciplinary-specific metadata sets that are consistent with items 1 and 2.
DC Principles

- “Core” across all knowledge domains
- No element required
- All elements repeatable
- 1:1 principle
DC encodings

- HTML `<meta>`
- XML
- RDF
- [Spreadsheets]
- [Databases]
Content/value standards for DC

- None required
- Some elements recommend a content or value standard as a best practice
  - Coverage
  - Date
  - Format
  - Language
  - Identifier
  - Relation
  - Source
  - Subject
  - Type
Some limitations of DC

- Can’t indicate a main title vs. other subordinate titles
- No method for specifying creator roles
- W3CDTF format can’t indicate date ranges or uncertainty
- Can’t by itself provide robust record relationships
Good times to use DC

- Cross-collection searching
- Cross-domain discovery
- Metadata sharing
- Describing some types of simple resources
- Metadata creation by novices
<table>
<thead>
<tr>
<th>Record format</th>
<th>DC</th>
<th>QDC</th>
<th>MARC</th>
<th>MARCXML</th>
<th>MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>record</td>
<td>record</td>
<td>record</td>
<td>record</td>
<td>record</td>
</tr>
<tr>
<td>XML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(X)HTML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field labels</td>
<td>Text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliance on</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AACR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common method of creation</td>
<td>By novices, by specialists, and by derivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Qualified Dublin Core (QDC)

- Adds some increased specificity to Unqualified Dublin Core
- Same governance structure as DC
- Same encodings as DC
- Same content/value standards as DC
- Listed in DMCI Terms
- Additional principles
  - Extensibility
  - Dumb-down principle
Types of DC qualifiers

- Additional elements
- Element refinements
- Encoding schemes
  - Vocabulary encoding schemes
  - Syntax encoding schemes
DC qualifier status

- Recommended
- Conforming
- Obsolete
- Registered
Limitations of QDC

- Widely misunderstood
- No method for specifying creator roles
- W3CDTF format can’t indicate date ranges or uncertainty
- Split across 3 XML schemas
- No encoding in XML officially endorsed by DCMI
Best times to use QDC

- More specificity needed than simple DC, but not a fundamentally different approach to description
- Want to share DC with others, but need a few extensions for your local environment
- Describing some types of simple resources
- Metadata creation by novices
<table>
<thead>
<tr>
<th>Record format</th>
<th>DC</th>
<th>QDC</th>
<th>MARC</th>
<th>MARCXML</th>
<th>MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[record]</td>
<td>[record]</td>
<td>[record]</td>
<td>[record]</td>
<td>[record]</td>
</tr>
<tr>
<td>XML</td>
<td>XML</td>
<td>XML</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF (X)HTML</td>
<td>RDF (X)HTML</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field labels</th>
<th>DC</th>
<th>QDC</th>
<th>MARC</th>
<th>MARCXML</th>
<th>MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[collection]</td>
<td>[collection]</td>
<td>[collection]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td>Text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reliance on AACR</th>
<th>DC</th>
<th>QDC</th>
<th>MARC</th>
<th>MARCXML</th>
<th>MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Common method of creation</th>
<th>DC</th>
<th>QDC</th>
<th>MARC</th>
<th>MARCXML</th>
<th>MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>By novices, by specialists, and by derivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MAchine Readable Cataloging (MARC)

- Format for the records in IUCAT and other OPACs
- Used for library metadata since 1960s
  - Adopted as national standard in 1971
  - Adopted as international standard in 1973
- Maintained by:
  - Network Development and MARC Standards Office at the Library of Congress
  - Standards and the Support Office at the National Library of Canada
More about MARC

- Actually a family of MARC standards throughout the world
  - U.S. & Canada use MARC21
- Structured as a binary interchange format
  - ANSI/NISO Z39.2
  - ISO 2709
- Field names
  - Numeric fields
  - Alphabetic subfields
Content/value standards for MARC

- None required by the format itself
- But US record creation practice relies heavily on:
  - AACR2r
  - ISBD
  - LCNAF
  - LCSH
Limitations of MARC

- Use of all its potential is time-consuming
- OPACs don’t make full use of all possible data
- OPACs virtually the only systems to use MARC data
- Requires highly-trained staff to create
- Local practice differs greatly
Good times to use MARC

- Integration with other records in OPAC
- Resources are like those traditionally found in library catalogs
- Maximum compatibility with other libraries is needed
- Have expert catalogers for metadata creation
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>XML</td>
<td>XML</td>
<td>ISO 2709</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF (X)HTML</td>
<td>RDF (X)HTML</td>
<td>[ANSI Z39.2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field labels</td>
<td>Text</td>
<td>Text</td>
<td>Numeric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliance on AACR</td>
<td>None</td>
<td>None</td>
<td>Strong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common method of creation</td>
<td>By novices, by specialists, and by derivation</td>
<td>By novices, by specialists, and by derivation</td>
<td>By specialists</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MARC in XML (MARCXML)

- Copies the exact structure of MARC21 in an XML syntax
  - Numeric fields
  - Alphabetic subfields
- Implicit assumption that content/value standards are the same as in MARC
Limitations of MARCXML

- Not appropriate for direct data entry
- Extremely verbose syntax
- Full content validation requires tools external to XML Schema conformance
Best times to use MARCXML

- As a transition format between a MARC record and another XML-encoded metadata format
- Materials lend themselves to library-type description
- Need more robustness than DC offers
- Want XML representation to store within larger digital object but need lossless conversion to MARC
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>XML</td>
<td>XML</td>
<td>ISO 2709</td>
<td>XML</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF (X)HTML</td>
<td>RDF (X)HTML</td>
<td>[ANSI Z39.2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XML</td>
<td>XML</td>
<td></td>
<td>XML</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDF (X)HTML</td>
<td>RDF (X)HTML</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td>Text</td>
<td>Numeric</td>
<td>Numeric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>None</td>
<td>Strong</td>
<td>Strong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>By novices, by specialists, and by derivation</td>
<td>By novices, by specialists, and by derivation</td>
<td>By specialists</td>
<td>By derivation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Metadata Object Description Schema (MODS)

- Developed and managed by the Library of Congress Network Development and MARC Standards Office
- First released for trial use June 2002
- MODS 3.0 released December 2003
- “Schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications.”
Differences between MODS and MARC

- MODS is “MARC-like” but intended to be simpler
- Textual tag names
- Encoded in XML
- Some specific changes
  - Some regrouping of elements
  - Removes some elements
  - Adds some elements
Content/value standards for MODS

- Many elements indicate a given content/value standard should be used
  - Generally follows MARC/AACR2/ISBD conventions
  - But not all enforced by the MODS XML schema
- Authority attribute available on many elements
Limitations of MODS

- No lossless round-trip conversion from and to MARC
- Still largely implemented by library community only
- Some semantics of MARC lost
Good times to use MODS

- Materials lend themselves to library-type description
- Want to reach both library and non-library audiences
- Need more robustness than DC offers
- Want XML representation to store within larger digital object
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>XML</td>
<td>XML</td>
<td>ISO 2709</td>
<td>XML</td>
<td>XML</td>
<td></td>
</tr>
<tr>
<td>RDF (X)HTML</td>
<td>RDF (X)HTML</td>
<td>[ANSI Z39.2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td>Text</td>
<td>Numeric</td>
<td>Numeric</td>
<td>Text</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>None</td>
<td>Strong</td>
<td>Strong</td>
<td>Implied</td>
<td></td>
</tr>
<tr>
<td>By novices, by specialists, and by derivation</td>
<td>By novices, by specialists, and by derivation</td>
<td>By specialists</td>
<td>By derivation</td>
<td>By specialists and by derivation</td>
<td></td>
</tr>
</tbody>
</table>
Mapping between metadata formats

- Also called “crosswalking”
- To create “views” of metadata for specific purposes
- Mapping from robust format to more general format is common
- Mapping from general format to more robust format is ineffective
Types of mapping logic

- Mapping the complete contents of one field to another
- Splitting multiple values in a single local field into multiple fields in the target schema
- Translating anomalous local practices into a more generally useful value
- Splitting data in one field into two or more fields
- Transforming data values
- Boilerplate values to include in output schema
Common mapping pitfalls

- Cramming in too much information
- Leaving in trailing punctuation
- Missing context of records
- Meaningless placeholder data

ALWAYS remember the purpose of the metadata you are creating!
No, really, which one do I pick?

- It depends. Sorry.
- Be as robust as you can afford
- Plan for future uses of the metadata you create
- Leverage existing expertise as much as possible
- Focus on content and value standards as much as possible
More information

- Dublin Core
  - DC Element Set version 1.1
  - DCMI Metadata Terms
- MODS
- MARC
- MARCXML
Questions?

- Jenn Riley, Metadata Librarian, IU Digital Library Program: jenlrile@indiana.edu
- These presentation slides: <http://www.dlib.indiana.edu/~jenlrile/presentations/bbspr05/descMDBB/>