714: Metadata

Metadata Structure Standards: Dublin Core and Digital Resource Description

Margaret E.I. Kipp - kipp@uwm.edu
https://pantherfile.uwm.edu/kipp/public/courses/714
Characteristics of Metadata Schemas

- **Structure**
  - what information is included in the metadata (the data model) and how it is structured

- **Syntax**
  - how the metadata is encoded

- **Semantics**
  - the meaning of the various elements of the metadata
Metadata Standards

• Metadata schema
  – defines important elements
  – defines required elements
  – defines structures/encodings for element sets
  – defines or suggests value standards (controlled vocabularies)
  – defines or suggests best practices
Two Example Metadata Schemas

• Dublin Core
  – 15 main elements
  – encoded using HTML or XML
  – frequently extended
  – most elements are optional
  – http://dublincore.org/

• AACR2R + MARC
  – 8 main areas, implementation of ISBD + MARC encoding, numeric
  – each chapter covers a different format (e.g. monograph, maps, sound recordings)
Metadata Standards

• Standards for data **structures**
• Standards for data **content**
• Standards for data **values**
• Standards for data **encoding** and **exchange**
Data Structure Standards

• Metadata element sets (schemas)
  – Standards for data structures and semantics
  – Covers required and optional elements
  – Guides to best practices in creating elements sets
  – e.g. Dublin Core Metadata Element Set (DCMES)
    http://dublincore.org/documents/dces/
Data Content Standards

• Guide to metadata generation or cataloguing
  – Describes the content which should be placed in each metadata element
  – May also provide formatting instructions for arrangement or display
  – Examples:
    • AACR2 (Anglo-American Cataloguing Rules)
    • Cataloguing Cultural Objects (CCO)
    • Describing Archives: A Content Standard (DACS)
Data Value Standards

• Value encoding schemes
  – Includes controlled vocabularies (e.g. subject heading lists), term lists, classification schemes, thesauri and authority files
  – Examples:
    • Library of Congress Subject Headings (LCSH)
    • Art and Architecture Thesaurus (AAT)
    • ISO 3166 Country Codes
Data Exchange Standards

• Crosswalks and transfer formats for sharing metadata records between schemas
• May be included in the schema or created separately
• e.g. MARC, XML
• XML is a general markup format which can be extended and defined via a DTD (Document Type Definition) or an XSD (XML Schema Document)
Creating Metadata

• 3 steps:
  
• **description** (providing a description of the information package and any information necessary for use)

• **encoding** (providing the syntax for the metadata)

• **access** (providing access to the metadata)
Creating Metadata 2

• Select unit to be described:
  – Collection versus volume versus item

• Select important descriptive information from the intellectual content of the item

• Determine important format characteristics of the item

• Place this information in a standard format for easy retrieval
RDA and FRBR

- RDA (Resource Description and Access) update to AACR2
  - FRBR: Functional Requirements for Bibliographic Records
    - Item = particular copy of a book
    - Manifestation = the item in general as opposed to the single copy
    - Expression = a translation or adaptation of a work
    - Work = the intellectual content as developed by the author
  - [http://www.loc.gov/cds/downloads/FRBR.PDF](http://www.loc.gov/cds/downloads/FRBR.PDF)
FRBR Examples

• Sample work level descriptions (http://frbr.oclc.org/research/pages/):
  – http://frbr.oclc.org/research/pages/000021759.html
  – http://frbr.oclc.org/research/pages/000276976.html

• At what level should metadata be generated?

• Is the work-item distinction important for all instances of metadata creation?
AACR2: Anglo-American Cataloguing Rules 2nd revision

• Based on the ISBD (International Standard Bibliographic Description)
• 2 parts: Description and Access Points
• 8 basic elements for metadata
• 13 chapters, one for each type of format
• Mixes work and item information in record
• content standard, not an encoding standard
AACR2 and MARC

- AACR2 defines a structure standard and data standard (Part 1: Description) and defines rules for value standards (Part 2: Access Points)
- MARC (MAchine Readable Cataloguing) defines a data encoding and exchange standard
AACR2 Elements

• Area 1: Title and Statement of Responsibility (and format - General Material Designation)
• Area 2: Edition
• Area 3: Material specific details
• Area 4: Publication, distribution, etc.
• Area 5: Physical Description
• Area 6: Series
• Area 7: Notes
• Area 8: Standard number and terms of availability
AACR2 Formats

- 1. General
- 2. Books (Monographs)
- 3. Maps
- 4. Manuscripts
- 5. Printed music
- 6. Sound recordings
- 7. Moving images
- 8. Graphic Material
- 9. Electronic resources (software, data files, websites, etc.)
- 10. 3D Objects (realia, kits, toys)
- 11. Microforms
- 12. Continuing resources (serials, sets, etc.)
- 13. Analytics
Combining AACR2 and MARC for Metadata Creation

- 1. Description - AACR2
- 2. Name access - AACR2
- 3. Subject Analysis – various controlled vocabularies
- 4. Record formatting - MARC
- 5. Record organisation - MARC
Base Elements and Formatting

• AACR2 provides lists of required elements and a required format on three levels of complexity.

• e.g. Title proper / first statement of responsibility. -- Edition statement. -- Material (or type of publication) specific details. -- First publisher, etc., date of publication, etc. -- Physical description. -- Note(s). -- Standard number.
AACR2 Data Content and Values

• AACR2 Part 1 defines where to find the information and how to arrange/display it in the record
• AACR2 Part 2 defines which elements contain controlled terms and how to generate them (authority controlled author names, corporate names, and uniform titles etc.)
Sample AACR2 Record with Access Points


• 1. Libraries and people with disabilities--United States. 2. People with disabilities--Legal status, laws, etc. -- United States.

• I. Foos, Donald D. II. Pack, Nancy C. III. Jahoda, Gerald.

• 027.663

• Z711.92.H3 .H6 1992
AACR2 Fields for Electronic Resources

- AACR2 was developed before the web so there is limited support for electronic resources in the topical areas.
- Most information goes in a note field.
- URL: http://example.com/example.html.
- The GMD for an electronic resource is: [electronic resource].
MARC Elements

- 3 digit field code - names element (245)
- 2 indicator codes – format instructions (10)
- at least one subfield indicator – separates out subelements ($c)

- e.g.

- 245 10 $a Metadata / $c Marcia Lei Zeng and Jian Qin
MARC Record Example

- 245 00 How libraries must comply with the Americans with Disabilities Act (ADA) / $c compiled and edited by Donald D. Foos and Nancy C. Pack.
- 300 ## xxii, 168 p. ; $c 23 cm.
- 504 ## Includes bibliographical references and index.
Value Standards in MARC

- subject headings are 650 fields
- the number after 650 indicates which controlled vocabulary was used
- notice that this particular record has controlled vocabulary from 3 sources:
  - LCSH (Library of Congress Subject Headings)
  - MeSH (Medical Subject Headings)
  - Repertoire de vedettes-matiere (from Library and Archives of Canada)
Other MARC Fields of Interest

- Description begins in field 245 and ends in field 504 (notes).

- Subject/name access (value standards)
  - 1XX - controlled name access
  - 700 - names and titles based on AACR2,
  - 650 - subject headings
  - 050 - LCC and 082 - DDC

- Provenance and History of Record
  - 040 list of libraries that have modified record
    DLC = Library of Congress
MARC Fields for Electronic Resources

• 245 $h GMD

• 538 System Requirements
  – e.g. 538 ## $a Mode of Access: World Wide Web
    • system requirements or how the resource is accessed, more important when many electronic resources were on CD or via Telnet

• 856 Electronic Location and Access
  – e.g. 856 40 $u http://www.oclc.org$zAccess using login and password.
    • the URL plus display text
Commonalities in Library Cataloguing Codes

- Early cataloguing codes based on items not works
- Emphasis on display format as well as content (encoding added later)
- Importance of subject access (subject headings and classification)/multiple access points
- Importance of author, title, subject
- Emphasis on location of information (IR)
Encoding Example: Metadata for Dummies by Jane Q. Smith

- AACR2: Metadata for Dummies / by Jane Q. Smith

- AACR2/MARC:
  - 245 10 $a Metadata for Dummies / $c by Jane Q. Smith

- AACR2/MARCXML:

- `<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Metadata for Dummies /
  </subfield>
  <subfield code="c">by Jane Q. Smith</subfield>
</datafield>`
In Class Exercise: Creating an AACR2/MARC record

- Create an AACR2/MARC record using one of the citations from the following exercise:
Dublin Core

- intended to be a core set of metadata elements and encoding conventions designed to be used across platforms and different document repositories
  - 15 basic elements:
    - contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type
Dublin Core Metadata Initiative (DCMI)

• Dublin Core Semantics
  – DC metadata elements, extensions, refinements and controlled vocabularies

• DCMI Abstract Model
  – Shared data model which sets an underlying grammar for expressing DC element sets

• Application Profiles
  – Implementation agreements (as in AACR2, describes how to apply DC)
Basic Description with DC

- DC intended to contain only those elements necessary for the discovery of the resources
- 15 basic elements in DCMES 1.1
- Intended to minimise the effort required to create good quality metadata
- Few rules for format/content compared to AACR2
DC Elements

Content
- title
- description
- type
- subject
- source
- relation
- coverage

Intellectual Property
- creator
- publisher
- rights
- contributor

Instantiation
- date
- format
- identifier
- language
Term Name: format
URI: http://purl.org/dc/elements/1.1/format
Label: Format
Definition: The file format, physical medium, or dimensions of the resource.
Comment: Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].
References: [MIME] http://www.iana.org/assignments/media-types/
[From DCMES1.1 http://dublincore.org/documents/dces/]
DC's One to One Principle

• DC metadata describes one manifestation or version of a resource
  – e.g. Does not assume that a print book and audio book could be equivalent
• Distinguish between reproductions and original works
Dublin Core Refinements

• Refine an element by providing more specific detail about the data value given

• Previously known as qualifiers
  – e.g. DC.description.tableOfContents = a DC description element which contains a table of contents
DC Refinements 2

• Originally just qualifiers for an element
• In 2003 declared to be terms or subelements (DCMI Metadata Terms)
• Allows refinements to be used separately from elements
  – e.g. dc.description.tableOfContents
  – Or <dcterm:tableOfContents>
• http://dublincore.org/documents/dcmi-terms/
DC Refinements Example

- dc.identifier.citation Schmiesing, R.J. (Ed.). (2006). Ohio 4-H youth development research: Selected abstracts (4H RES NO. 1). Columbus, OH: The Ohio State University Extension

- <meta name="DCTERMS.bibliographicCitation" content="Schmiesing, R.J. (Ed.). (2006). Ohio 4-H youth development research: Selected abstracts (4H RES NO. 1). Columbus, OH: The Ohio State University Extension" xml:lang="en" />

- https://kb.osu.edu/dspace/handle/1811/24305?mode=full
<table>
<thead>
<tr>
<th>Element</th>
<th>Refinement</th>
<th>Encoding Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>alternative</td>
<td></td>
</tr>
<tr>
<td>creator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>subject</td>
<td></td>
<td>DDC; LCC; LCSH; NLM; UDC; MESH</td>
</tr>
<tr>
<td>description</td>
<td>tableOfContents; abstract</td>
<td></td>
</tr>
<tr>
<td>publisher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>contributor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>date</td>
<td>created; valid; issued; modified; dateCopyrighted; dateSubmitted; available; dateAccepted</td>
<td>W3CDTF</td>
</tr>
<tr>
<td>type</td>
<td></td>
<td>DCMIType</td>
</tr>
<tr>
<td>format</td>
<td>extent; medium</td>
<td>IMT</td>
</tr>
<tr>
<td>identifier</td>
<td>bibliographicCitation</td>
<td>URI</td>
</tr>
<tr>
<td>source</td>
<td></td>
<td>URI</td>
</tr>
<tr>
<td>language</td>
<td></td>
<td>RFC4646; ISO639-2; ISO639-3; RFC1766; RFC3066</td>
</tr>
<tr>
<td>relation</td>
<td>IsVersionOf; isReplacedBy; isRequiredBy; isPartOf; isReferencedBy; isFormatOf; conformsTo; hasVersion; replaces; requires; hasParts; references; hasFormat</td>
<td></td>
</tr>
<tr>
<td>coverage</td>
<td>spatial; temporal</td>
<td>Box; Point; Period; ISO3166; TGN; W3CDTF</td>
</tr>
<tr>
<td>rights</td>
<td>accessRights; license</td>
<td></td>
</tr>
</tbody>
</table>
DCMI Namespaces

• What is a namespace?
  – a space in which each term is unique and can only be used to mean one thing

• DCMI metadata terms have a unique URI within the DCMI namespaces
  – URI = Uniform Resource Identifier, URIs, like URLs must be unique

• DCMI namespaces
  – http://purl.org/dc/elements/1.1/ (15 DC elements)
  – http://purl.org/dc/terms/ (DCMI terms)
  – http://purl.org/dc/dcmitype/ (DCMI type vocab)
DCMI Namespaces 2

• Title is defined at:
  – http://purl.org/dc/elements/1.1/title (DCMI 15)
  or
  – http://purl.org/dc/terms/title (DCMI terms)

• This URI describes the element and serves as a unique identifier for the element

• The URI can be referenced in an XML DTD (Data Type Definition) to define the element
DCMI Value Encoding Schemes

- Value encoding schemes control the values that are entered into metadata elements
- Vocabulary encoding schemes
  - ISO639-2 (language codes), LCSH, DDC, etc.
- Syntax encoding schemes
  - W3CDTF (W3C Encoding Rules for Date and Time)
DC and XML

- Dublin Core is generally encoded in HTML or XML

- HTML has predefined tags for encoding
  - e.g. http://eprints.rclis.org/handle/10760/8720

- XML is the most popular general encoding method because you can define the tags in a DTD or an XSD
  - e.g. https://pantherfile.uwm.edu/kipp/public/courses/714/metadataexamples/dcinxml-onlinethesis.xml
HTML and XML Tags

- `<HTML></HTML>`
- A tag is an element in the HTML or XML schema
- Tags are markup
- define how to display (HTML) or what is contained (XML and some HTML) in the tag
  - e.g. `<title>This is a title</title>`
  - `<meta name="DC.title" content="This is a title."/>`
DC XML Tags

• DC XML tags consist of the qualified names of DC elements
  – e.g. <dc:title>This is a title</dc:title>
  – The content of the metadata statement is in between the tags

• http://dublincore.org/documents/dc-xml-guidelines/
DC Terms XML Tags

• DC Terms (includes original 15 elements and refinements) XML tags also consist of the qualified names of DC elements

• e.g. `<dcterms:tableOfContents>Moving Targets in Professional Learning 32</dcterms:tableOfContents>`

• or
  `<dc:description.tableOfContents>Moving Targets in Professional Learning 32</dc:description.tableOfContents>`
DC XML DTD

- early XML relied on a DTD (Data Type Definition) to define its tags and explain how they should be displayed
- XSD (XML Schema) is a newer standard for defining tags, etc.
  - [http://dublincore.org/schemas/xmls/](http://dublincore.org/schemas/xmls/)
Encoding Example 2

• Dublin Core:
  – dc.title: Metadata for Dummies
  – dc.creator: Jane Q. Smith

• Dublin Core/HTML:
  – <meta name="DC.title" content="Metadata for Dummies" /> <meta name="DC.creator" content="Jane Q. Smith" />

• Dublin Core/XML:
In Class Exercise: Creating a DC Record (in HTML or XML)

• Create a DC record encoded in HTML or XML using one of the citations from the following exercise:

• Example:
  - in HTML: <meta name="DC.title" content="This is a title."/>
  - in XML: <dc:title>This is a title</dc:title>
  - see https://pantherfile.uwm.edu/kipp/public/courses/714/metadataexamples/ for examples