Digital Project Planning & Management Basics

Optional Unit: Specific metadata standards and applications overviews

Addendum to session 4
Session Objectives

- Understand standards for
  - Metadata elements
  - Data value standards
  - Data content standards and

- Learn about metadata standards developed by specific communities

- Evaluate the efficacy of the standard for a specific community, their strengths and weaknesses

- Explore the adoption of non-traditional standards by libraries
Session Outline

- Introduction to basic concepts
- Description of community specific metadata schemes
- Description of specific structural metadata and syntaxes
Questions to Ask When Selecting a Metadata Standard

- What type of material will be digitized?
- How much information is available?
- Is there a Community of practice developed for this resource type(s)?
- What is the purpose of digital project?
- Did your “Needs Assessment” elicit who will be the audience and how they would use the content?
- Are there pre-existing digital projects with which this one needs to function?
- What Systems options are available?
Metadata Standards in a Resource Grid

- **Books**
  - MARC, DC
  - ONIX, MPEG

- **Journals**
  - MARC, DC
  - ONIX, MPEG

- **Special Collections**
  - MARC, METS, EAD, DC, TEI

- **Freely-accessible web resources**
  - DC
  - DDI, IEEE/LOM, FGDC, EAD, TEI, SCORM

- **Institutional assets**
  - Books
  - Journals
  - Newspapers
  - Government docs
  - Audiovisual
  - Maps
  - Scores
  - Special collections
    - Rare books
    - Local/Historical Newspapers
    - Local history materials
    - Archives & manuscripts
    - Theses & dissertations

- **Unique-ness**
  - High
  - Low

- **Stewardship**
  - DC

(Based on Lorcan Dempsey Presentation)
Metadata Standards

- Schemas (a.k.a. ‘Element Sets’)
  - Set of semantic properties, in this context used to describe resources
  - Not the same as “XML schemas” (which has a very precise meaning)

- Syntaxes
  - The structural wrapping around the semantics
  - Essential for moving information around
Content Standards

- AACR2 functions as the content standard for traditional cataloging
- RDA (the successor to AACR2) aspires to be the content standard for non-MARC metadata
- DACS (Describing Archives: a Content Standard)
- CCO (cataloging Cultural Objects) new standard developed by visual arts and cultural heritage community
- Best practices, Guidelines, Data dictionaries-- less formal content standards
Value Standards

- Library of Congress Subject Headings
- Art and Architecture Thesaurus
- Thesaurus of Geographical Names
Some Example Schemas

- Dublin Core (http://dublincore.org)
  - Simple and Qualified
- MODS (www.loc.gov/standards/mods/)
- VRA 4.0 (http://www.vraweb.org/projects/vracore4/index.html)
- IEEE-LOM (http://ltsc.ieee.org/wg12/)
- ONIX (http://www.editeur.org/onix.html)
- EAD (http://www.loc.gov/ead/)
- TEI (http://www.tei-c.org/)
Dublin Core: Simple

- Fifteen elements; one namespace
- Controlled vocabulary values may be expressed, but not the sources of the values
- Minimal standard for OAI-PMH
- Used also as:
  - core element set in some other schemas
  - switching vocabulary for more complex schemas
Dublin Core Metadata Initiative (DCMI) Origins

- 2nd W3C Conference Chicago (October 1994)
  - Conversations at this conference led to the first meeting at OCLC in Dublin Ohio, hence its name
  - Combination of IT and Librarians

- Workshops began in 1995
  - March 1995, NCSA/OCLC workshop in Dublin, Ohio
  - Identified the need for author generated metadata, a “core”: of common elements to describe Web content to help discovery
Mission of the DCMI (Original)

“The mission of the Dublin Core Metadata Initiative (DCMI) is to make it easier to find resources using the Internet through the following activities:

- Developing metadata standards for resource discovery across domains
- Defining frameworks for the interoperation of metadata sets
- Facilitating the development of community- or domain-specific metadata sets that work within these frameworks”

DCMES Characteristics

- Simplicity
- Supports resource discovery
- All elements are optional/repeatable
- No order of elements prescribed
- Extensible* / Refined*
- Interdisciplinary/International
- Semantic interoperability
Value

- International and cross-domain
- Increase efficiency of the discovery/retrieval of digital objects
- Provide a framework of elements which will aid the management of information
- Promote collaboration of cultural/educational information as shared “social capital”
DCMES Principles

- 1:1
- Dumb Down
- Appropriate Values

http://dublincore.org/documents/usageguide/glossary.shtml
Dublin Core Metadata Element Set (DCMES) 1996

The 15 Dublin Core elements can be divided into three categories:

<table>
<thead>
<tr>
<th>CONTENT</th>
<th>INTELLECTUAL PROPERTY</th>
<th>INSTANTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Creator</td>
<td>Date</td>
</tr>
<tr>
<td>Description</td>
<td>Contributor</td>
<td>Language</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></td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ex.: Simple Dublin Core

<metadata>
  <dc:title>Cataloging cultural objects,</dc:title>
  <dc:contributor>Baca, Murtha.</dc:contributor>
  <dc:contributor>Harpring, Patricia.</dc:contributor>
  <dc:subject>Information organization</dc:subject>
  <dc:subject>Metadata</dc:subject>
  <dc:subject>Cultural property--Documentation</dc:subject>
  <dc:subject>CC135.C37 2006</dc:subject>
  <dc:subject>363.6</dc:subject>
  <dc:date>2006</dc:date>
  <dc:format>396 p.</dc:format>
  <dc:type>Text</dc:type>
  <dc:language>en</dc:language>
  <dc:publisher>ALA Editions</dc:publisher>
</metadata>
Extensible: Lego Blocks

- Extensible architecture
  - Spectrum of simple to more complex
  - DCMES may be used with other metadata element sets
  - Lego™ Metaphor: Modular building blocks used to develop application profiles of mixed metadata
- Leverage existing thesauri, classification systems, ontologies, local vocabularies

Dublin Core: Qualified

- ‘Qualified’ includes element refinements and encoding schemes
  - More specific properties
  - Two namespaces
  - Explicit vocabularies
- Additional elements, including ‘Audience,’ ‘InstructionalMethod,’ ‘RightsHolder’ and ‘Provenance’
## Qualified Dublin Core

<table>
<thead>
<tr>
<th>Elements</th>
<th>Element Refinements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Identifier</td>
<td>Abstract</td>
</tr>
<tr>
<td>2. Title</td>
<td>Access rights</td>
</tr>
<tr>
<td>3. Creator</td>
<td>Alternative</td>
</tr>
<tr>
<td>4. Contributor</td>
<td>Audience</td>
</tr>
<tr>
<td>5. Publisher</td>
<td>Available</td>
</tr>
<tr>
<td>6. Subject</td>
<td>Bibliographic citation</td>
</tr>
<tr>
<td>7. Description</td>
<td>Conforms to</td>
</tr>
<tr>
<td>8. Coverage</td>
<td>Created</td>
</tr>
<tr>
<td>9. Format</td>
<td>Date accepted</td>
</tr>
<tr>
<td>10. Type</td>
<td>Date copyrighted</td>
</tr>
<tr>
<td>11. Date</td>
<td>Date submitted</td>
</tr>
<tr>
<td>12. Relation</td>
<td>Education level</td>
</tr>
<tr>
<td>13. Source</td>
<td>Extent</td>
</tr>
<tr>
<td>14. Rights</td>
<td>Has format</td>
</tr>
<tr>
<td>15. Language</td>
<td>Has part</td>
</tr>
<tr>
<td></td>
<td>Has version</td>
</tr>
<tr>
<td></td>
<td>Is format of</td>
</tr>
<tr>
<td></td>
<td>Is part of</td>
</tr>
<tr>
<td></td>
<td>Is referenced by</td>
</tr>
<tr>
<td></td>
<td>Is replaced by</td>
</tr>
<tr>
<td></td>
<td>Is required by</td>
</tr>
<tr>
<td></td>
<td>Issued</td>
</tr>
<tr>
<td></td>
<td>Is version of</td>
</tr>
<tr>
<td></td>
<td>License</td>
</tr>
<tr>
<td></td>
<td>Mediator</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Modified</td>
</tr>
<tr>
<td></td>
<td>Provenance</td>
</tr>
<tr>
<td></td>
<td>References</td>
</tr>
<tr>
<td></td>
<td>Replaces</td>
</tr>
<tr>
<td></td>
<td>Requires</td>
</tr>
<tr>
<td></td>
<td>Rights holder</td>
</tr>
<tr>
<td></td>
<td>Spatial</td>
</tr>
<tr>
<td></td>
<td>Table of contents</td>
</tr>
<tr>
<td></td>
<td>Temporal</td>
</tr>
<tr>
<td></td>
<td>Valid</td>
</tr>
</tbody>
</table>
More Dublin Core
“Refinements”

<table>
<thead>
<tr>
<th>Encodings</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box</td>
<td>Collection</td>
</tr>
<tr>
<td>DCMIType</td>
<td>Dataset</td>
</tr>
<tr>
<td>DDC</td>
<td>Event</td>
</tr>
<tr>
<td>IMT</td>
<td>Image</td>
</tr>
<tr>
<td>ISO3166</td>
<td>Interactive Resource</td>
</tr>
<tr>
<td>ISO639-2</td>
<td>Moving Image</td>
</tr>
<tr>
<td>LCC</td>
<td>Physical Object</td>
</tr>
<tr>
<td>LCSH</td>
<td>Service</td>
</tr>
<tr>
<td>MESH</td>
<td>Software</td>
</tr>
<tr>
<td>Period</td>
<td>Sound</td>
</tr>
<tr>
<td>Point</td>
<td>Sound</td>
</tr>
<tr>
<td>RFC1766</td>
<td>Still Image</td>
</tr>
<tr>
<td>RFC3066</td>
<td>Text</td>
</tr>
<tr>
<td>TGN</td>
<td></td>
</tr>
<tr>
<td>UDC</td>
<td></td>
</tr>
<tr>
<td>URI</td>
<td></td>
</tr>
<tr>
<td>W3CTDF</td>
<td></td>
</tr>
</tbody>
</table>
Ex.: Qualified Dublin Core

<metadata>
  <dc:title xml:lang="en">Cataloging cultural objects.</dc:title>
  <dc:contributor>Baca, Murtha.</dc:contributor>
  <dc:contributor>Harpring, Patricia.</dc:contributor>
  <dc:subject xsi:type="LCSH">Information organization</dc:subject>
  <dc:subject xsi:type="LCSH">Metadata</dc:subject>
  <dc:subject xsi:type="LCSH">Cultural property--Documentation</dc:subject>
  <dc:subject xsi:type="LCC">CC135.C37 2006</dc:subject>
  <dc:subject xsi:type="DDC">363.3</dc:subject>
  <dc:date xsi:type="W3CDTF">2006</dc:date>
  <dcterms:extent>396 p.</dcterms:extent>
  <dc:type xsi:type="DCMIType">Text</dc:type>
  <dc:language xsi:type="RFC3066">en</dc:language>
  <dc:publisher>ALA Editions</dc:publisher>
  <dcterms:audience>Catalogers</dcterms:audience>
</metadata>
Lego Model replaced by RDF

Combining element sets using the Resource Description Framework (RDF), Semantic Web

Container

- Package
  - Dublin Core

- Package
  - MARC record

- Package
  - Indirect Reference

URI

- Package
  - Terms and Conditions
Advantages of Dublin Core

- Less rigorous content rules
- Easier to train and implement
- Allows OAI harvesting of metadata
- Supported by digital library products:
  - ContentDM
  - Encompass
  - MetaSource
Disadvantages to Dublin Core

- Lack of granularity may not support specific community needs
- Lack of granularity makes its role as a switching language between standards limited
- No fields are required and lack of consistent training can hamper interoperability
What is MODS?

Descriptive metadata standard

- Initiative of Network Development and MARC Standards Office at LC
- A derivative of MARC21
  - Documentation refers to MARC definitions for most properties
  - Descriptive metadata encoded in an XML schema
  - Uses textual rather than numeric tags
- Originally designed for library applications, but may be used for others
- Uses XML Schema (METS)

http://www.loc.gov/standards/mods/
XML (Extensible Markup Language) is the markup for the Web

Library community need for a element set simpler but compatible with MARC that could be transmitted in XML

A standardized framework for holding and exchanging metadata: analogous to the MARC record, for re-use of pre-existing information

Designed for complex digital library objects

Dublin Core not sufficient; e.g., need to express role of creator

Provide a more explicit means of expressing different categories of dates in machine-readable forms
elements

- Title Info
- Name
- Type of resource
- Genre
- Origin Info
- Language
- Physical description
- Abstract
- Table of contents
- Target audience
- Note

- Subject
- Classification
- Related item
- Identifier
- Location
- Access conditions
- Extension
- Record Info

Root elements:

- mods (A single MODS record)
- modsCollection (A collection of MODS records)
Election 2002 Web Archive Record

Title: FranUlmer.com -- Home Page

Alternative Title: Fran Ulmer, Democratic Party candidate for Governor, Alaska, 2002.

Name: Ulmer, Fran

Abstract: Web site promoting the candidacy of Fran Ulmer, Democratic Party candidate for Governor, Alaska, 2002. Includes candidate biography, issue position statements, campaign contact information, privacy policy and campaign news press releases. Site features enable visitors to sign up for campaign email list, volunteer, make campaign contributions and follow links to other internet locations.

Date Captured: July 2, 2002 - November 18, 2002

Archived Site

Subjects: Elections--Alaska
Governors--Alaska--Election
Democratic Party (AK)

Language: English

Genre: Web Site

Access Condition: None

Active Site: http://www.franulmer.com/

Collection Title: Election 2002 Web Archive

The Library of Congress
March 4, 2003

Close window to return to search page

Contact Us
Fields used in Minerva project

- Title
- Alternative title
- Name (structured form)
- Abstract
- Date captured
- Genre (value always “Web site”)
- Physical description (file formats)
- Identifier (base URL)
- Language
- Access conditions/rights management
- Subject (keyword or LCSH if possible)
Advantages of MODS:

- Uses language-based tags; fully uses Unicode character set
- Allows the aggregation of multilingual records
- Elements generally inherit semantics of MARC but does not assume the use of any specific rules for description
- Element set is more compatible with existing descriptions than ONIX or Dublin Core
- Elements particularly applicable to digital resources
- XML schema allows for flexibility and availability of freely available software tools
Disadvantages of MODS

- Library-centric
- Not widely adopted by other libraries or other communities
Ex.: MODS

<titleInfo>
  <title>Cataloging cultural objects. /</title>
</titleInfo>

<name type="personal">
  <namePart type="family">Baca,</namePart>
  <namePart type="given">Murtha),</namePart>
  <namePart type="date">1951-</namePart>
  <role>
    <roleTerm type="text">editor</roleTerm>
  </role>
</name>

<name type="personal">
  <namePart type="family">Harpring,</namePart>
  <namePart type="given">Patricia. </namePart>
  <role>
    <roleTerm type="text">editor</roleTerm>
  </role>
</name>
More MODS

<typeOfResource>text</typeOfResource>
<genre authority="marc">book</genre>
<originInfo>
  <place>
    <placeTerm authority="marccountry" type="code">ilu</placeTerm>
  </place>
  <place>
    <placeTerm type="text">Chicago</placeTerm>
  </place>
  <publisher>ALA Editions</publisher>
  <dateIssued>2006</dateIssued>
  <issuance>monographic</issuance>
</originInfo>
<language>
  <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
</language>
VRA Core Categories for Visual Resources

- Developed by the Visual Resources Association, the VRA Standards Committee
- Designed specifically for visual resources
- Viewed as a means to share cataloging of visual materials
- Provides access to digitized images and their description
VRA Metadata Elements

- Based on CDWA for category definitions and recommendations for controlled vocabulary
- Two types of elements
  - Work
  - Images
- Like DC, all fields are repeatable
- Unlike DC, all are mandatory if applicable
VRA 4.0 Elements

- Work, Collection or Image
- Work Type
- Title
- Measurements
- Material
- Technique
- Agent
- Date
- Subject
- Relation
- Location REFID
- Text REF
- Style/Period
- Agent.Culture / Cultural Context
- Description
- Source
- Rights
- Inscription
- State Edition
VRA Data Values

- LCSH
- AAT
- TGN
- ULAN
Online Information Exchange (ONIX)

Designed by publishing industry (American Association of Publishers) to exchange information about “books” with wholesalers, retail, e-tail booksellers.

- Standard for data exchange
- Richer information for online bookstores
ONIX Integrated with MARC Records?

CC:DA Task on ONIX International charge with reviewing the standard and assessing the impact if integrated

http://www.ala.org/alcts/organization/ccs/ccda/tf-onix1.html
Comparison of ONIX & MARC

- ONIX has finer granularity than MARC
- Fields can be mapped from ONIX into UNIMARC, but cannot be reconverted
- Each application contains fields that are relevant to only themselves
- ONIX records provide enriching information: reviews, abstracts, TOC, prizes won, credentials of authors
ONIX/MARC Crosswalks

- ONIX (1.0) to UNIMARC Crosswalk developed by Library of Congress
  [http://lcweb.loc.gov/marc/onix2marc.html](http://lcweb.loc.gov/marc/onix2marc.html)

- Mapping by Bob Pearson (OCLC)

- Report by Alan Danskin
  [http://bic.org.uk/reporton.doc](http://bic.org.uk/reporton.doc)
ONIX Metadata Standard

Allows two levels of description:

- **Level 2:**
  - 235 elements of information in 24 categories
  - Requires XML DTD

- **Level 1:**
  - Not all the categories, 82 elements
  - Does not require XML DTD
ONIX for Books

- Originally devised to simplify the provision of book product information to online retailers (name stood for ONline Information eXchange)

- First version flat XML, second version included hierarchy and elements repeated within ‘composites’

- Maintained by Editeur, with the the Book Industry Study Group (New York) and Book Industry Communication (London)

- Includes marketing and shipping oriented information: book jacket blurb and photos, full size and weight info, etc.
Ex.: ONIX

<Title>
<TitleType>01</TitleType>
<TitleText textcase = “02”>British English, A to Zed</TitleText>
</Title>

<Contributor>
<SequenceNumber>1</SequenceNumber>
<ContributorRole>A01</ContributorRole>
<PersonNameInverted>Schur, Norman W</PersonNameInverted>
<BiographicalNote>A Harvard graduate in Latin and Italian literature, Norman Schur attended the University of Rome and the Sorbonne before returning to the United States to study law at Harvard and Columbia Law Schools. Now retired from legal practise, Mr Schur is a fluent speaker and writer of both British and American English.</BiographicalNote>
</Contributor>
BRITISH ENGLISH, A TO ZED is the thoroughly updated, revised, and expanded third edition of Norman Schur’s highly acclaimed transatlantic dictionary for English speakers. First published as BRITISH SELF-TAUGHT and then as ENGLISH ENGLISH, this collection of Briticisms for Americans, and Americanisms for the British, is a scholarly yet witty lexicon, combining definitions with commentary on the most frequently used and some lesser known words and phrases. Highly readable, it’s a snip of a book, and one that sorts out – through comments in American – the “Queen’s English” – confounding as it may seem.

Norman Schur is without doubt the outstanding authority on the similarities and differences between British and American English. BRITISH ENGLISH, A TO ZED attests not only to his expertise, but also to his undiminished powers to inform, amuse and entertain. – Laurence Urdang, Editor, VERBATIM, The Language Quarterly, Spring 1988
Ex.: ONIX

BRITISH ENGLISH, A TO ZED is the thoroughly updated, revised, and expanded third edition of Norman Schur’s highly acclaimed transatlantic dictionary for English speakers. First published as BRITISH SELF-TAUGHT and then as ENGLISH ENGLISH, this collection of Briticisms for Americans, and Americanisms for the British, is a scholarly yet witty lexicon, combining definitions with commentary on the most frequently used and some lesser known words and phrases. Highly readable, it’s a snip of a book, and one that sorts out – through comments in American – the “Queen’s English” – confounding as it may seem.

Norman Schur is without doubt the outstanding authority on the similarities and differences between British and American English. BRITISH ENGLISH, A TO ZED attests not only to his expertise, but also to his undiminished powers to inform, amuse and entertain. – Laurence Urdang, Editor, VERBATIM, The Language Quarterly, Spring 1988

Main Desc.

Review
EAD -- Encoded Archival Description

Encoded Archival Description (EAD)

Official EAD Version 2002 Web Site

The EAD Document Type Definition (DTD) is a standard for encoding archival finding aids using the Standard Generalized Markup Language (SGML). The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress (LC) in partnership with the Society of American Archivists.

General Information

- Development of the EAD DTD
- Design Principles for Enhancements to EAD
- EAD Listserv: How to subscribe | List message archive
- EAD Sites on the World Wide Web
- Other Sites Related to EAD
- EAD Working Group Members

http://www.loc.gov/ead/
Learning Object Metadata

- An array of related standards for description of ‘learning objects’ or ‘learning resources’
- Most based on efforts of the IEEE LTSC (Institute of Electrical and Electronics Engineers Learning Technology Standards Committee) and the IMS Global Learning Consortium, inc.
- Tends to be very complex with few implementations outside of government and industry
- One well-documented implementation is CanCore
The Library of Congress' Network Development and MARC Standards Office, in partnership with the NISO Technical Metadata for Digital Still Images Standards Committee and other interested experts, is developing an XML schema for a set of technical data elements required to manage digital image collections. The schema provides a format for interchange and/or storage of the data specified in the NISO Draft Standard Data Dictionary: Technical Metadata for Digital Still Images (Version 1.2). This schema is currently in draft status and is being referred to as "NISO Metadata for Images in XML (NISO MIX)". MIX is expressed using the XML schema language of the World Wide Web Consortium. MIX is maintained for NISO by the Network Development and MARC Standards Office of the Library of Congress with input from users.

This is a DRAFT for review and trial use: Please send comments on draft 0.2 to the MIX Listserv (described below)

XML schema for a set of technical data elements required to manage digital image collections

http://www.loc.gov/standards/mix/
Welcome to the TEI Website

Initially launched in 1987, the TEI is an international and interdisciplinary standard that helps libraries, museums, publishers, and individual scholars represent all kinds of literary and linguistic texts for online research and teaching, using an encoding scheme that is maximally expressive and minimally obsolescent.

- **All about the TEI Consortium**: describes the organization and constitution of the TEI Consortium
- **How to participate**: provides information on how projects, institutions, and individuals can play an active part in development and maintenance of the standard
- **The TEI Guidelines**: the chief deliverable of the TEI project: detailed recommendations for the encoding of all kinds of textual material of all kinds in all languages from all times
- **TEI Tutorials**: introductory and advanced teaching materials, presentations, and user case studies
- **TEI History**: archive of TEI publications and working papers
- **Projects using TEI**: pointers to live TEI applications and systems worldwide
- **Members only area**: links to current TEI activities, draft documents, discussion papers from workgroups, etc. (access restricted to current members only. (Forgotten your password? [contact us for a reminder])
- **Just the FAQs**: quick answers to frequently asked questions about the TEI
- **TEI Software**: pointers to TEI-specific and generic free software for exploiting the TEI scheme

http://www.tei-c.org/