Cataloging for the 21st Century -- Course 2

Metadata Standards & Applications

Student Manual

Original course design by
Diane I. Hillmann
Cornell University Library
Revised by Rebecca Guenther and
Allene Hayes, Library of Congress

For
The Library of Congress

And the
Association for Library Collections & Technical Services

Library of Congress R Cataloging Distribution Service
Washington, DC
March 2008

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Cataloging for the 21st Century
Course 2: Metadata Standards and Applications

Schedule

Day 1	Sessions / Topics
8:00-9:00	**Registration. Breakfast**
9:00-9:15	**Getting started; Introductions, Orientation and Background**
9:15-10:15	**1. Introduction to Digital Libraries and Metadata** • Discuss similarities and differences between traditional and digital libraries • Understand how the environment where metadata is developing is different from the library automation environment • Explore different types and functions of metadata (administrative, technical, administrative, etc.) Exercise: Examine three digital library instances, discuss differences in user approach and experience, and look for examples of metadata use
10:15-10:30	Break
10:30-12:00	**2. Descriptive Metadata Standards** • Understand the categories of descriptive metadata standards (e.g., data content standards, data value standards, data structure standards, relationship models) • Learn about the various descriptive metadata standards and the communities that use them • Evaluate the efficacy of a standard for a particular community • Understand how relationship models are used Exercise: Create a brief descriptive metadata record using the standard assigned.
12:00-1:30	Lunch
1:30-3:00	**3. Technical and Administrative Metadata Standards** • Understand the different types of administrative metadata • Learn about the metadata needed for supporting digital preservation activities • Understand the importance of technical, structural and rights metadata in digital libraries Exercise: Provide technical metadata for the same resource used in the descriptive exercise.
3:00-3:15	Break
3:15-4:30	**4. Metadata Syntaxes and Containers** • Overview of syntaxes, including HTML/XHTML, XML, RDF/XML • Overview of containers, including METS and MPEG-21 DID • Discover how container formats are used for managing digital resources and their metadata Exercise: Encode a simple resource description in Dublin Core, MARC, and MODS using XML
4:30	**Conclusion of Day 1**
Day 2	Sessions / Topics
9:00-10:30	**5. Applying Metadata Standards: Application Profiles** • Learn about the concept and use of application profiles, including examining specific types (e.g., Dublin Core Application Profiles, METS Profiles) • Learn how different metadata standards are used together in digital library applications • Understand questions for determining the appropriate metadata standard to use
10:30-10:45	Break
11:00-11:45	**6. Applying Metadata Standards: Controlled Vocabularies and Quality Issues** • Understand how different controlled vocabularies are used in metadata • Introduction to vocabulary encodings, including MARC 21 and the Simple Knowledge Organization

	System (SKOS), an evolving encoding for thesauri • Investigate metadata quality issues
11:45-1:15	Lunch
1:15-2:30	**7. Approaches to Metadata Creation, Storage and Retrieval in the Digital Environment** • Understand the differences between traditional vs. digital library metadata creation • Examine metadata creation management models • Investigate standards and methods of information retrieval and discovery Exercise: Using example websites, determine the underlying models for each
2:30-2:45	Break
2:45-4:00	**8. Metadata Interoperability and Crosswalks** • Understand interoperability protocols (OAI-PMH for harvesting, OpenURL for references) • Introduction to crosswalking and mapping Exercise: Examine the MARC to DC crosswalk and the DC to MARC crosswalk, determining where data loss occurs

Metadata Standards and Applications

Outline

1. Introduction to Digital Libraries and Metadata

- Discuss similarities and differences between traditional and digital libraries
- Understand how the environment where metadata is developing is different from the library automation environment
- Explore different types and functions of metadata (administrative, technical, administrative, etc.)

Exercise: Examine three digital library instances, discuss differences in user approach and experience, and look for examples of metadata use

2. Descriptive Metadata Standards

- Understand the categories of descriptive metadata standards (e.g., data content standards, data value standards, data structure standards, relationship models)
- Learn about the various descriptive metadata standards and the communities that use them
- Evaluate the efficacy of a standard for a particular community
- Understand how relationship models are used

Exercise: Create a brief descriptive metadata record using the standard assigned.

3. Technical and Administrative Metadata Standards

- Understand the different types of administrative metadata
- Learn about the metadata needed for supporting digital preservation activities
- Understand the importance of technical, structural and rights metadata in digital libraries

Exercise: Provide technical metadata for the same resource used in the descriptive exercise.

4. Metadata Syntaxes and Containers

- Overview of syntaxes, including HTML/XHTML, XML, RDF/XML
- Overview of containers, including METS and MPEG-21 DID
- Discover how container formats are used for managing digital resources and their metadata

Exercise: Encode a simple resource description in Dublin Core, MARC, and MODS using XML

5. Applying Metadata Standards: Application Profiles

- Learn about the concept and use of application profiles, including examining specific types (e.g., Dublin Core Application Profiles, METS Profiles)
- Learn how different metadata standards are used together in digital library applications
- Understand questions for determining the appropriate metadata standard to use

6. Applying Metadata Standards: Controlled Vocabularies and Quality Issues

- Understand how different controlled vocabularies are used in metadata
- Introduction to vocabulary encodings, including MARC 21 and the Simple Knowledge Organization System (SKOS), an evolving encoding for thesauri
- Investigate metadata quality issues

7. Approaches to Metadata Creation, Storage and Retrieval in the Digital Environment

- Understand the differences between traditional vs. digital library metadata creation
- Examine metadata creation management models
- Investigate standards and methods of information retrieval and discovery

Exercise: Using example websites, determine the underlying models for each

8. Metadata Interoperability and Crosswalks

- Understand interoperability protocols (OAI-PMH for harvesting, OpenURLs for references)
- Introduction to crosswalking and mapping

Exercise: Examine the MARC to DC crosswalk and the DC to MARC crosswalk, determining where data loss occurs

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Metadata Standards and Applications

Introduction: Background, Goals, and Course Outline

Original course design by Diane I. Hillmann, Cornell University Library
Revised by Rebecca Guenther and Allene Hayes, Library of Congress

Version 2, March, 2008

Cataloging for the 21st Century

Background for this course:

- The first of five courses developed as as part of:
  - Bibliographic Control of Web Resources: A Library of Congress Action Plan
  - Action Item 5.3: Continuing Education (CE)
  - Continuing Education Implementation Group (CEIG)
- See course Bibliography for citations

Cataloging for the 21st Century: The five CE course components

- 1. MARC/AACR2 cataloging of Internet resources
- 2. Overview of basic concepts for 21st century for bibliographic control, including specific metadata standards and applications
- 3. Thesaurus design principles, building a controlled vocabulary, and examination of selected controlled vocabularies
- 4. Digital library design, with a project-based component
- 5. Philosophy and approach to asset management for the 21st century, including evaluation skills

Cataloging for the 21st Century: CE Course Series Objectives

- To equip catalogers to deal with new types of resources and to recognize their unique characteristics
- To equip catalogers to evaluate competing approaches to and standards for providing access to resources
- To equip catalogers to think creatively and work collaboratively with others inside and outside their home institutions
- To ensure that catalogers have a broad enough understanding of the current environment to be able to make their local efforts compatible and interoperable with other efforts
- To prepare catalogers to be comfortable with ambiguity and being less than perfect
- To enable practicing catalogers to put themselves into the emerging digital information environment and to continue to play a significant role in shaping library services

Goals for this Course

- Understand similarities and differences between traditional and digital libraries
- Explore different types and functions of metadata (administrative, descriptive, technical, etc.)
- Understand metadata standards: schemas, data content standards, and data value standards
- Learn how various metadata standards are applied in digital projects, including use of application profiles
- Understand how different controlled vocabularies are used in digital libraries
- Approaches to metadata creation, storage and retrieval
- Learn about metadata interoperability and crosswalks

Course objectives

- Increase catalogers' understanding of metadata for digital resources
- Evaluate competing approaches and standards for managing and providing access to resources
- Enable catalogers to think creatively and work collaboratively
- Increase understanding of current environment to allow for compatibility among applications
- Increase flexibility in utilizing different kinds of metadata standards
- Allow catalogers to use expertise to contribute to the emerging digital information environment

Outline of this course

- Session 1. Introduction to digital libraries and metadata
- Session 2. Descriptive metadata standards
  - Data content standards, data value standards, data structure standards, relationship models
  - Specific descriptive metadata formats

Outline of this course cont.

- Session 3. Technical and administrative metadata standards
- Session 4. Syntaxes and containers for metadata
- Session 5. Applying metadata standards in the digital library
  - Profiling
  - Controlled vocabularies
  - Using metadata for managing and navigating
  - Quality issues

Outline of this course cont.

- Session 6. Metadata creation, storage and retrieval
- Session 7. Metadata interoperability and crosswalks

1. Introduction to Digital Libraries and Metadata

Metadata Standards and Applications Workshop

1

Goals of Session

- Understand similarities and differences between traditional and digital libraries focusing on metadata
- Explore different types and functions of metadata (descriptive, administrative, structural, etc.)

2

Traditional vs. Digital Libraries

Traditional library characteristics

Digital library characteristics?

3

What is a digital library?

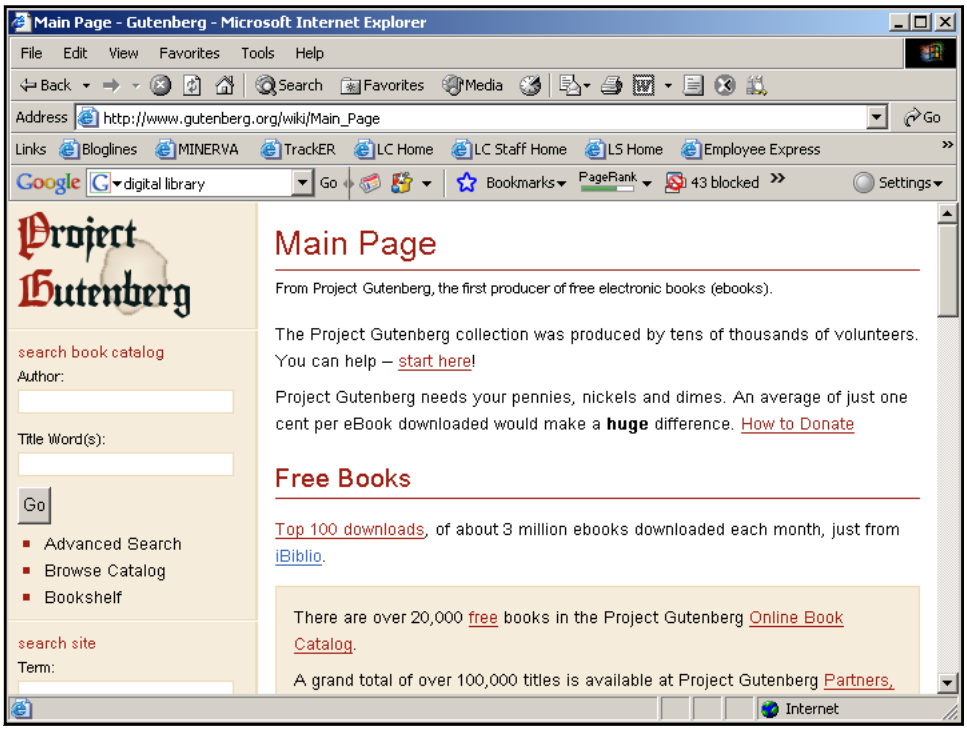
- a library in which collections are stored in digital formats and accessed by computers. The digital content may be stored locally, or accessed remotely via computer networks.
- a type of information retrieval system.

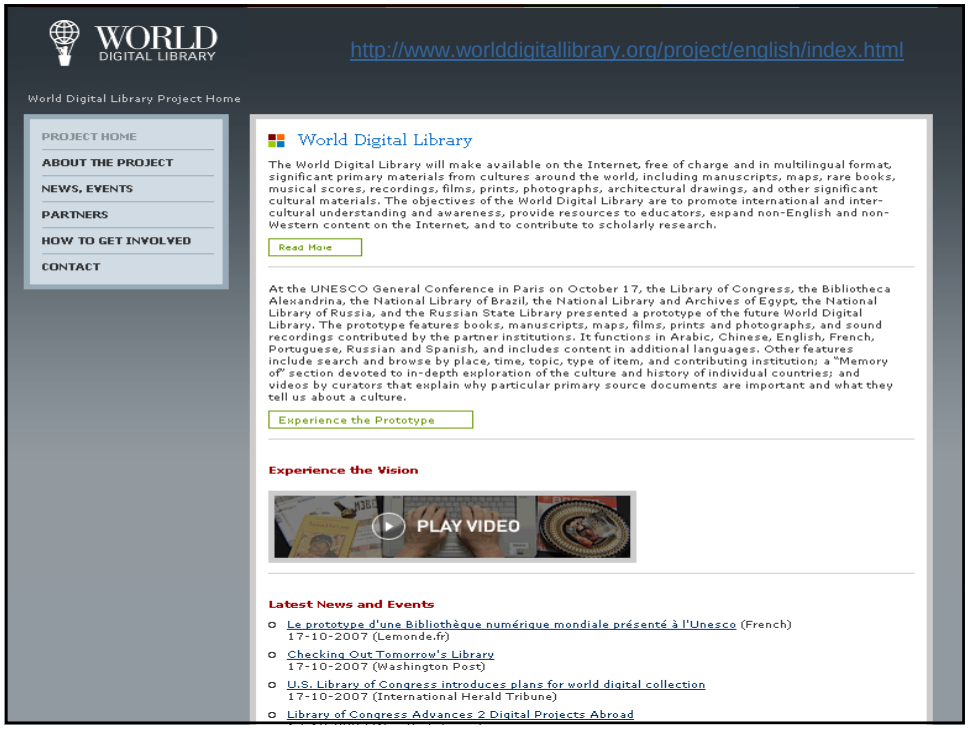
4

Digital Library Federation (DLF)

- "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."
- http://www.diglib.org/

5





How does the environment affect the creation of metadata?

8

Traditional Libraries

- Firm commitment to standards
  - Specifications for metadata content (e.g., AACR2)
  - Specifications for metadata encoding (e.g., MARC)
  - A variety of syntactical bindings available
- Agreements on quality expectations
- Tradition of sharing, facilitated by bibliographic utilities
- Available documentation and training

9

Digital Libraries

- No dominant content standard
- A variety of “formats” (or “schemas” or “element sets”)
- Some emerging “federated” agreements, mostly in the world of digital libraries attached to traditional libraries
- Variable quality expectations
- Emerging basis for sharing (OAI-PMH)
- Some documentation and training is becoming available

10

Environmental Factors

- Differences:
  - *Players*: New world of metadata not necessarily led by librarians
  - *Goals*: Competition for users critical for sustainability
  - *Resources*: No real basis for understanding non-technical needs (including metadata creation and maintenance)
  - Many levels of content responsibility (or none)

11

Environmental Factors

- Similarities
  - It's about discovery (and access, and use and meeting user needs)
  - Pressure for fast, cheap and "good enough" (also rich, scalable, and re-usable--is that a contradiction?)
  - Wide variety of materials and services
  - Maintenance needs often overlooked

12

What *IS* Metadata?

- Some possibilities:
  - Data about data (or data about resources)
  - A management tool
  - Computer-processible, human-interpretable information about digital and non-digital objects

13

"In moving from dispersed digital collections to interoperable digital libraries, the most important activity we need to focus on is standards… most important is the wide variety of metadata standards [including] descriptive metadata… administrative metadata…, structural metadata, and terms and conditions metadata…"

Howard Besser, NYU

14

Metadata standards in digital libraries

- Interoperability and object exchange requires the use of established standards
- Many digital objects are complex and are comprised of multiple files
- XML is the de-facto standard for metadata descriptions on the Internet
- Complex digital objects require many more forms of metadata than analog for their management and use
  - Descriptive
  - Administrative
    - Technical
    - Digital provenance/events
    - Rights/Terms and conditions
  - Structural

15

Types of metadata

- Descriptive
- Administrative
  - Technical
  - Digital provenance
  - Rights/Access
- Preservation
- Structural
- Meta-metadata
- Other?

16

Descriptive Metadata

- Title, author, human-readable description of a resource
- Subject or topical information
- Genre and format of the resource
- Relationships with other resources (version, parent/child, etc.)

17

Administrative Metadata

- Metadata to manage the object
  - Technical metadata: technical characteristics about the object
  - Digital provenance metadata: actions that have been performed on the object
  - Rights metadata: information about access and use of the object

18

Rights/Access Metadata

- Where is the resource? Is it in a place open to me?
- Are there restrictions on the use of the resource?
- What can I do with this resource?

19

Preservation Metadata

- Designed to ensure that the information the resource contains remains accessible to users over a long period of time
- Records details about format migration and data refreshment
- Tracks versions used for different kinds of access and display
- Allows a variety of approaches to the problem of maintaining resources over time

20

Structural Metadata

- No single standard or best practice governs structural metadata creation
- Ties the components of a complex or compound resource together and makes the whole usable
- Enables flexible and local approaches to presentation and navigation
- Various approaches to sharing structural metadata exist (METS perhaps the best known)

21

Meta-metadata

- Metadata about the metadata
  - Who created this information?
  - When was it created?
  - When were links last checked?
  - Other update transactions?
- May be a component of some metadata schemes

22

Functions of Metadata

Discover resources	Manage documents	Control IP Rights
Identify versions	Certify authenticity	Indicate status
Mark content structure	Situate geospatially	Describe processes

23

Cataloging and Metadata

- Cataloging early form of desriptive metadata
  - Underlying models for cataloging based on AACR2 and MARC 21
  - Some new metadata models are emerging (e.g, DC Abstract Model and RDA in development)

24

One BIG Difference ...

- Catalogers most often are attempting to fit new items into an already existing world of materials--
  - The structure already exists, as do the rules for describing
- Metadata practitioners are generally working with aggregated "stuff," attempting to find a way to make it accessible
  - Involves broad understanding, ability to work with others to make decisions that work for whole projects or domains

25

**Thanks to Marty Kurth for these insights*

Exercise

- Examine the digital library sites below, and be prepared to discuss differences in user approach and experience. Look for how metadata is used.
  - Alsos: Digital Library for Nuclear Issues (http://alsos.wlu.edu/default.aspx)
  - CSUN Oviatt Library: Digital Collections (http://library.csun.edu/Collections/SCA/digicoll.html)
  - Birdsource (http://www.birdsource.org/)

26

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

2. Specific metadata standards: descriptive

Metadata Standards and Applications Workshop

1

Session 2 Objectives

- Understand the categories of descriptive metadata standards (e.g., data content standards vs. data value standards)
- Learn about the various descriptive metadata standards in terms of the community that developed them
- Evaluate the efficacy of the standard for a specific community, their strengths and weaknesses

2

Outline of Session 2: descriptive metadata

- Types of descriptive metadata standards (e.g. element sets, content standards)
- Specific descriptive metadata standards (e.g. MARC, DC, MODS, EAD...)

3

Descriptive metadata

- Most standardized and well understood type of metadata
- Major focus of library catalog
- Increased number of descriptive metadata standards for different needs and communities
- Importance for resource discovery
- May support various user tasks

4

Aspects of descriptive metadata

- Data content standards (e.g., rules: AACR2R/RDA, CCO)
- Data value standards (e.g., values/controlled vocabularies: LCNAF, LCSH, MeSH, AAT)
- Data structure standards (e.g., formats/schemes: DC, MODS, MARC 21, VRA Core)
- Relationship models
- Data exchange/syntax standards (e.g. MARC 21 (ISO 2709), MARCXML, DC/RDF or DC/XML)

5

Content Standards: Rules

- **AACR2** functions as the content standard for traditional cataloging
- **RDA** (*Resource Description and Access*) is the successor to AACR2 that aspires to be independent of a particular syntax
- **DACS** (*Describing Archives: a Content Standard*)
- **CCO** (*Cataloging Cultural Objects*) new standard developed by visual arts and cultural heritage community
- **CSDGM** (*Content Standards for Digital Geospatial Metadata)*
- Best practices, Guidelines, policies-- less formal content standards

6

Content Standards: Value Standards/Controlled Vocabularies

- Examples of thesauri
  - Library of Congress Subject Headings
  - Art and Architecture Thesaurus
  - Thesaurus of Geographical Names
- Examples of value lists
  - ISO 639-2 Language codes
  - MARC Geographic Area codes
  - Other enumerated lists (e.g. MARC/008 lists)
  - Dublin Core Resource Types

7

Data structure standards (element sets and formats)

- Facilitates database creation and record retrieval
- Flexibility because not tied to a particular syntax
- May provide a minimum of agreed upon elements that facilitate record sharing and minimal consistency
- Different user communities develop their own standard data element sets
- May differ in complexity and granularity of fields
- Some data element sets become formats/schemes by adding rules such as repeatability, controlled vocabularies used, etc.

8

Key descriptive metadata element sets/formats

- MARC 21
- MODS
- Dublin Core
- ONIX
- EAD
- Special resource types
  - CDWA/VRA Core
  - IEEE-LOM
  - PBCore
  - IPTC Core

9

What is MARC 21?

- A syntax defined by an international standard and was developed in the late 60s
- As a syntax it has 2 expressions:
  - Classic MARC (MARC 2709)
  - MARCXML
- A data element set defined by content designation and semantics
- Institutions do not store “MARC 21”, as it is a communications format
- Many data elements are defined by external content rules; a common misperception is that it is tied to AACR2

10

MARC 21 Scope

- **Bibliographic Data**
  - books, serials, computer files, maps, music, visual materials, mixed material
- **Holdings Data**
  - physical holdings, digital access, location
- **Authority Data**
  - names, titles, name/title combinations, subjects, series
- **Classification Data**
  - classification numbers, associated captions, hierarchies
- **Community Information**
  - events, programs, services, people, organizations

11

MARC 21 implementation

- National formats were once common and there were different flavors of MARC
- Now most have harmonized with MARC 21 (e.g. CANMARC, UKMARC, MAB)
- Billions of records world wide
- Integrated library systems that support MARC bibliographic, authority and holdings format
- Wide sharing of records for 30+ years
- OCLC is a major source of MARC records

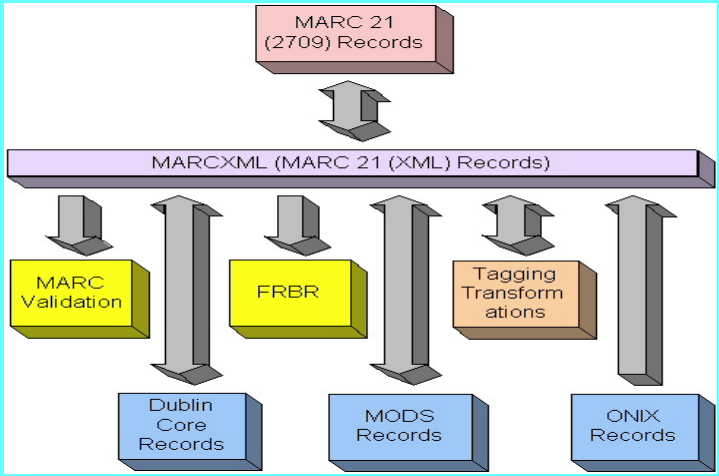
12

Streamlining MARC 21 into the future

- Take advantage of XML
  - Establish standard MARC 21 in an XML structure
  - Take advantage of freely available XML tools
- Develop simpler (but compatible) alternatives
  - MODS
- Allow for interoperability with different XML metadata schemas
  - Assemble coordinated set of tools
- Provide continuity with current data
  - Provide flexible transition options

13

MARC 21 evolution to XML



MARC 21 in XML – MARCXML

- MARCXML record
  - XML exact equivalent of MARC (2709) record
  - Lossless/roundtrip conversion to/from MARC 21 record
  - Simple flexible XML schema, no need to change when MARC 21 changes
  - Presentations using XML stylesheets
  - LC provides converters (open source)
- http://www.loc.gov/standards/marcxml

15

Example: MARC and MARCXML

- Music record in MARC
- Music record in MARCXML

16

What is MODS?

- Metadata Object Description Schema
- An XML descriptive metadata standard
- A derivative of MARC
  - Uses language based tags
  - Contains a subset of MARC data elements
  - Repackages elements to eliminate redundancies
- MODS does not assume the use of any specific rules for description
- Element set is particularly applicable to digital resources

17

MODS high-level 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
- Part
- Extension
- Record Info

18

Advantages of MODS

- Element set is compatible with existing descriptions in large library databases
- Element set is richer than Dublin Core but simpler than full MARC
- Language tags are more user-friendly than MARC numeric tags
- Hierarchy allows for rich description, especially of complex digital objects
- Rich description that works well with hierarchical METS objects

19

Uses of MODS

- Extension schema to METS
  - Rich description works well with hierarchical METS objects
- To represent metadata for harvesting (OAI)
  - Language based tags are more user friendly
- As a specified XML format for SRU
- As a core element set for convergence between MARC and non-MARC XML descriptions
- For original resource description in XML syntax that is simpler than full MARC

20

Example: MODS

- Music record in MODS

21

Status of MODS

- Open listserv collaboration of possible implementers, LC coordinated (1st half 2002)
- First comment and use period: 2nd half 2002
- Now in MODS version 3.3
- Registration approved by National Information Standards Organization (NISO)
- Companion for authority metadata (MADS) in version 1.0
- Endorsed as METS extension schema for descMD
- Many expose records as MODS in OAI

22

A selection of MODS projects

- LC uses of MODS
  - LC web archives
  - Digital library METS projects
- University of Chicago Library
  - Chopin early editions
  - Finding aid discovery
- Digital Library Federation Aquifer initiative
- National Library of Australia
  - MusicAustralia: MODS as exchange format between National Library of Australia and ScreenSoundAustralia
  - Australian national bibliographic database metadata project
- See: MODS Implementation registry
  http://www.loc.gov/mods/registry.php

23

What is MADS?

- Metadata Authority Description Schema
- A companion to MODS for authority data using XML
- Defines a subset of MARC authority elements using language-based tags
- Elements have same definitions as equivalent MODS
- Metadata about people, organizations, events, subjects, time periods, genres, geographics, occupations

24

MADS elements

- **authority**
  - name
  - titleInfo
  - topic
  - temporal
  - genre
  - geographic
  - hierarchicalGeographic
  - occupation
- **related**
  - same subelements
- **variant**
  - same subelements
- **note**
- **affiliation**
- **url**
- **identifier**
- **fieldOfActivity**
- **extension**
- **recordInfo**

25

Uses of MADS

- As an XML format for information about people, organizations, titles, events, places, concepts
  - To expose library metadata in authority files
  - To allow for linking to an authoritative form and fuller description of the entity from a MODS record
- For a simpler authority record than full MARC 21 authorities
- To integrate bibliographic/authority information for presentation

26

Examples

- person
- organization
- title
- topic
- genre
- geographic

27

Some MADS implementations

- Irish Virtual Research Library and Archive Repository Prototype
- Perseus Digital Library (Tufts)
- Mark Twain Papers (University of California)
- Library of Congress/National Library of Egypt

28

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

29

Dublin Core Metadata Element Set (DCMES) 1996

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

CONTENT	INTELLECTUAL PROPERTY	INSTANTATION
Title	Creator	Date
Description	Contributor	Language
Subject	Publisher	Identifier
Relation	Rights	Format
Source		
Coverage		
Type		

30

Advantages

- 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

31

Dublin Core Characteristics

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

32

Dublin Core metadata

- Original 15 included in DC simple
- Elements defined subsequently and all refinements/encoding schemes under dcterms (qualified)
- DCMI Type values for high level resource types
- Simple DC widely implemented and required for metadata harvesting using OAIPMH (until current version)
- Application profiles developing to document usage and additional elements needed
- http://dublincore.org

33

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'

34

Ex.: Simple Dublin Core

```
<metadata>
   <dc:title>3 Viennese arias: for soprano, obbligato clarinet in B flat, and
piano.</dc:title>
   <dc:contributor>Lawson, Colin (Colin James)</dc:contributor>
   <dc:contributor>Bononcini, Giovanni, 1670-1747.</dc:contributor>
<dc:contributor>Joseph I, Holy Roman Emperor, 1678-1711.</dc:contributor>
   <dc:subject>Operas--Excerpts, Arranged--Scores and parts</dc:subject>
   <dc:subject>Songs (High voice) with instrumental ensemble--Scores and
parts</dc:subject>
   <dc:subject>M1506 .A14 1984</dc:subject>
   <dc:subject></dc:subject>
   <dc:subject></dc:subject>
   <dc:date>1984</dc:date>
   <dc:format>1 score (12 p.) + 2 parts ; 31 cm.</dc:format>
   <dc:type>Sound</dc:type>
   <dc:identifier>85753651</dc:identifier>
   <dc:language>it</dc:language>
   <dc:language>en</dc:language>
   <dc:publisher>Nova Music</dc:publisher></metadata>
```

35

Ex.: Qualified Dublin Core

```
<metadata>
   <dc:title xml:lang="en">3 Viennese arias: for soprano, obbligato clarinet in B
flat, and piano.</dc:title>
   <dc:contributor>Lawson, Colin (Colin James)</dc:contributor>
   <dc:contributor>Bononcini, Giovanni, 1670-1747.</dc:contributor>
<dc:contributor>Joseph I, Holy Roman Emperor, 1678-1711.</dc:contributor>
   <dc:subject xsitype="LCSH">Operas--Excerpts, Arranged--Scores and
parts</dc:subject>
   <dc:subject xsitype="LCSH">Songs (High voice) with instrumental
ensemble--Scores and parts</dc:subject>
   <dc:subject xsitype="LCC">M1506 .A14 1984</dc:subject>
   <dc:date xsitype="W3CDTF">1984</dc:date>
 <dcterms:extent>1 score (12 p.) + 2 parts ; 31 cm.</dcterms:extent>
 <dc:type xsitype="DCMIType">Sound</dc:type>
   <dc:identifier>85753651</dc:identifier>
   <dc:language xsitype="RFC3066">it</dc:language>
   <dc:language xsitype="RFC3066">en</dc:language>
   <dc:publisher>Nova Music</dc:publisher>
</metadata>
```

36

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.

37

ONIX scheme

- Assigns textual element names as well as short alphanumeric tags
- Forms are identical in their functionality
- ONIX has extensive codelists
- Institutions could receive ONIX records from publishers and use an ONIX to MARC (or other metadata scheme) conversion

38

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>

39

ONIX

Main Desc.

<othertext>
<d102>01**</d102>**
<d104>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 A… the "Queen's English" – confounding as it may seem.**</d1**
</otherte
<othertext
<d102>08**</d102>**
<d104>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 **</d104>**
</othertext>

Review

40

Encoded Archival Description (EAD)

- Standard for electronic encoding of finding aids for archival and manuscript collections
- Expressed as an SGML/XML DTD
- Supports archival descriptive practices and standards
- Supports discovery, exchange and use of data
- Developed and maintained by Society of American Archivists; LC hosts the website

41

EAD, continued

- Based on the needs of the archival community
- Good at describing blocks of information, poor at providing granular information
- Some uptake by museum community
- Not a content standard
- EAC is a companion for information about creators of archival material
- Example: http://purl.dlib.indiana.edu/iudl/findingaids/lilly/InU-Li-VAA1292

42

Benefits of an EAD finding aid

- Documents the interrelated descriptive information of an archival finding aid
- Preserves the hierarchical relationships existing between levels of description
- Represents descriptive information that is inherited by one hierarchical level from another
- Supports element-specific indexing and retrieval of descriptive information

43

VRA Core

- Developed by the Visual Resources Association's Data Standards Committee
- Metadata element set for descriptions of work of visual culture
- Includes hierarchical structure
- Currently in version 4.0
- XML schema has been established for record sharing
- Data value standards may come from CCO (for content rules) and thesauri (e.g. TGM, AAT)

44

Work, Collection or Image

- work, collection or image
- agent
- culturalContext
- date
- description
- inscription
- location
- Material
- Measurements
- relation
- rights
- source
- stateEdition
- stylePeriod
- subject
- technique
- textRef
- title
- workType

45

Advantages of VRA

- Allows description of original and digital object
- Level of granularity greater than Dublin Core, less than MARC and supports specific discipline
- Now content rules have been developed (CCO)

46

VRA examples

- Lindesfarne gospels (manuscript)
- Chanel coat (3-dimensional object)
  - XML
  - Display

47

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

48

Ex.: CanCore

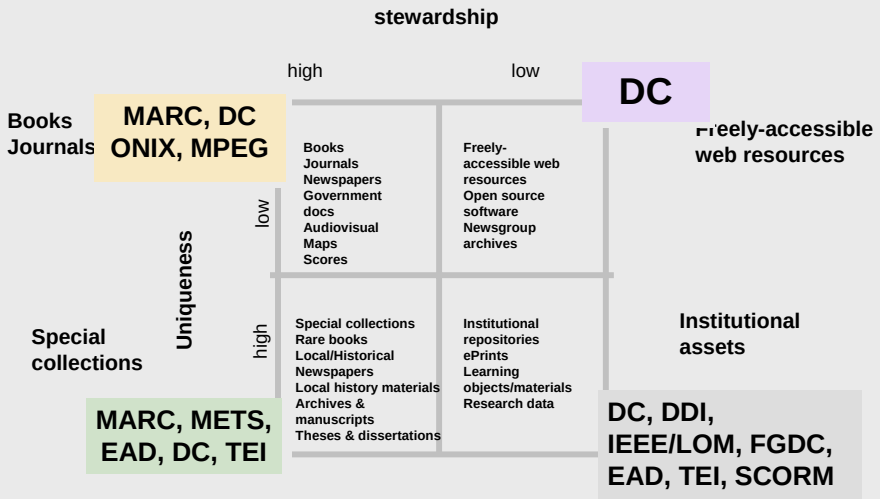
```
<learningResourceType>
  <source>LOMv1.0</source>
  <value>narrative text</value>
</learningResourceType>
<learningResourceType>
  <source>GEM Resource Type Controlled Vocabulary
http://www.geminfo.org/Workbench/Metadata/Vocab_Type.html
  </source>
  <value>educator's guide</value>
</learningResourceType>
<learningResourceType>
  <source>LOMv1.0</source>
  <value>narrative text</value>
</learningResourceType>
<learningResourceType>
  <source>EdNA Curriculum
  http://www.edna.edu.au/edna/go/cache/offonce/pid/621
  </source>
  <value>training package</value>
</learningResourceType>
```

Note name & URL

Note name & URL

49

Metadata Standards in a Resource Grid



Lorcan Dempsey cited in Stuart Weibel. Presentation **State of the Dublin Core Metadata Initiative** Göttingen August 11, 2003

50

Modeling metadata: why use models?

- To understand what entities you are dealing with
- To understand what metadata are relevant to which entities
- To understand relationships between different entities
- To organize your metadata to make it more predictable (and be able to use automated tools)

51

Descriptive metadata models

- Conceptual models for bibliographic and authority data
  - Functional Requirements for Bibliographic Records (FRBR)
  - Functional Requirement for Authority Data (FRAD)
- Dublin Core Abstract Model (DCAM)
- Some other models:
  - CIDOC Conceptual Reference Model (emerged from museum community)
  - INDECS (for intellectual property rights)
- There are many conceptual models intended for different purposes

52

Bibliographic relationships (pre-FRBR)

- Tillett's Taxonomy (1987)
  - Equivalence
  - Derivative
  - Descriptive
  - Whole-part
  - Accompanying
  - Sequential
  - Shared-characteristic

53

Bibliographic relationships in MARC/MODS

- MARC Linking entry fields
- MARC relationships by specific encoding format
  - Authority vs bibliographic vs holdings
- MODS relationships
  - relatedItem types
  - Relationship to METS document

54

FRBR (1996)

- IFLA Study Group on the Functional Requirements for Bibliographic Records
- Focused on the bibliographic record rather than the catalog
- Used an entity relationship model, rather than descriptive analysis without a structural model
- Broader in scope than previous studies

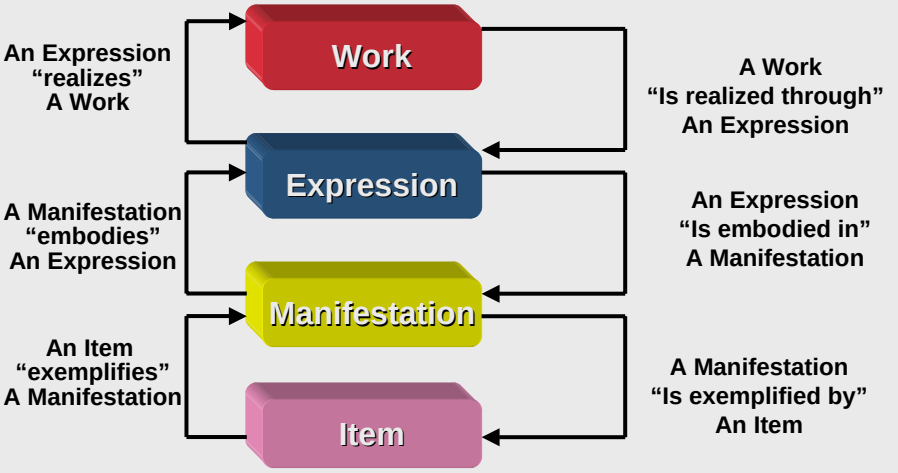
55

FRBR Entities

- Bibliographic entities: *works, expressions, manifestations, items*
- Responsible parties: *persons, corporate bodies*
- Subject entities: *concepts, objects, events, places*

56

Group 1 Entities and Relationships

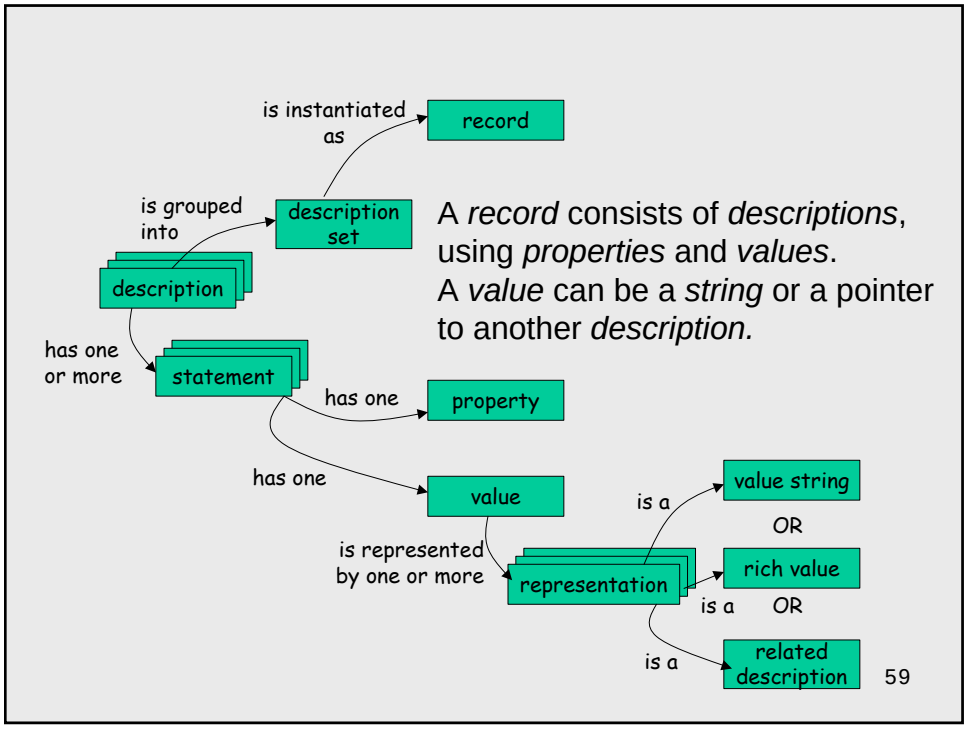


57

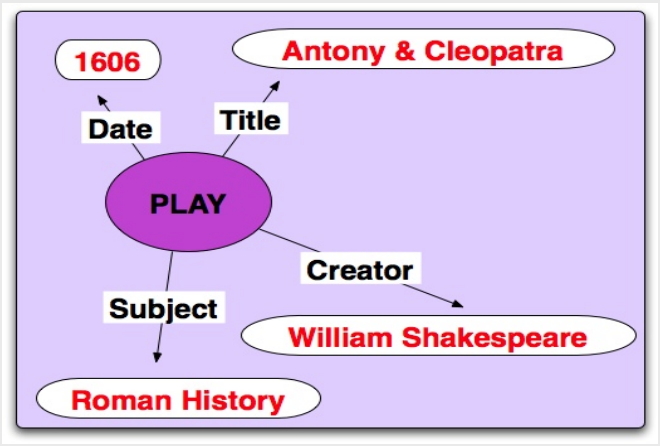
DC Abstract Model

- Reaffirms the One-to-One Principle
- Defines 'statement' as the atomic level
- Distinguishes between "description" and "description set":
  - **Description**: "A description is made up of one or more statements about one, and only one, resource."
  - **Description Set**: "A description set is a set of one or more descriptions about one or more resources."
- Work is being done to understand RDA in terms of the DCAM

58



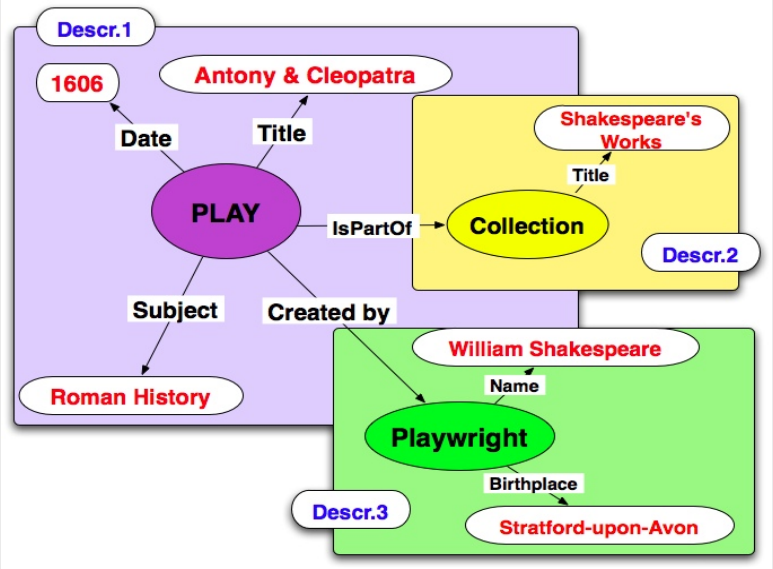
Basic model: Resource with properties



A Play has the title “Antony and Cleopatra,” was written in 1606 by William Shakespeare, and is about “Roman history”

60

... related to other Resources



61

An Exercise

- *Each group will be given a printout of a digital object*
- *Create a brief metadata record based on the standard assigned to your group*
- *Take notes about the issues and decisions made*
- *Appoint a spokesperson to present the metadata record created & the issues involved (5-10 minutes)*
- *62*

62

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Metadata Standards & Applications
Metadata Template
Exercise: Session 2, Descriptive Metadata

- Each group will be given a printout of a digital object
- Create a brief metadata record based on the standard assigned to your group
- Take notes about the issues and decisions made
- Appoint a spokesperson to present the metadata record created & the issues involved (2-5 minutes)

Provide participants with guides from metadata schemas:
(Note: Course materials provide "Using Dublin Core – The Elements")

Dublin Core Metadata Element Set Version 1.1
http://www.dublincore.org/documents/dces/ (provided)
Using Dublin Core—The Elements
http://dublincore.org/documents/usageguide/elements.shtml

MODS outline—version 3.3
http://www.loc.gov/standards/mods/mods-outline.html (provided)
MODS userguide:
http://www.loc.gov/standards/mods/v3/mods-userguide.html

VRA Core 4.0 Element Outline
http://www.vraweb.org/projects/vracore4/VRA_Core4_Outline.pdf (provided)
VRA Core 4.0 Element Description and Tagging Examples
http://www.vraweb.org/projects/vracore4/VRA_Core4_Element_Description.pdf

Provide participants with printout of different digital objects (selected by trainer).
Two examples:

(A) America's Pinch Hit March (sheet music from Baseball Sheet Music collection, LC)
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200033287/default.html

(B) Photograph of Katherine Dunham at the Theâtre de Paris in 1948.
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200003731/default.html

Participants will provide descriptive metadata in the standard assigned with the following elements (element names may vary according to scheme):

Title/subtitle
Creator/name (with role defined if possible)
Type of resource
Publication/origin information with place, publisher, date
Physical description
Subject

Use the Metadata Template below:

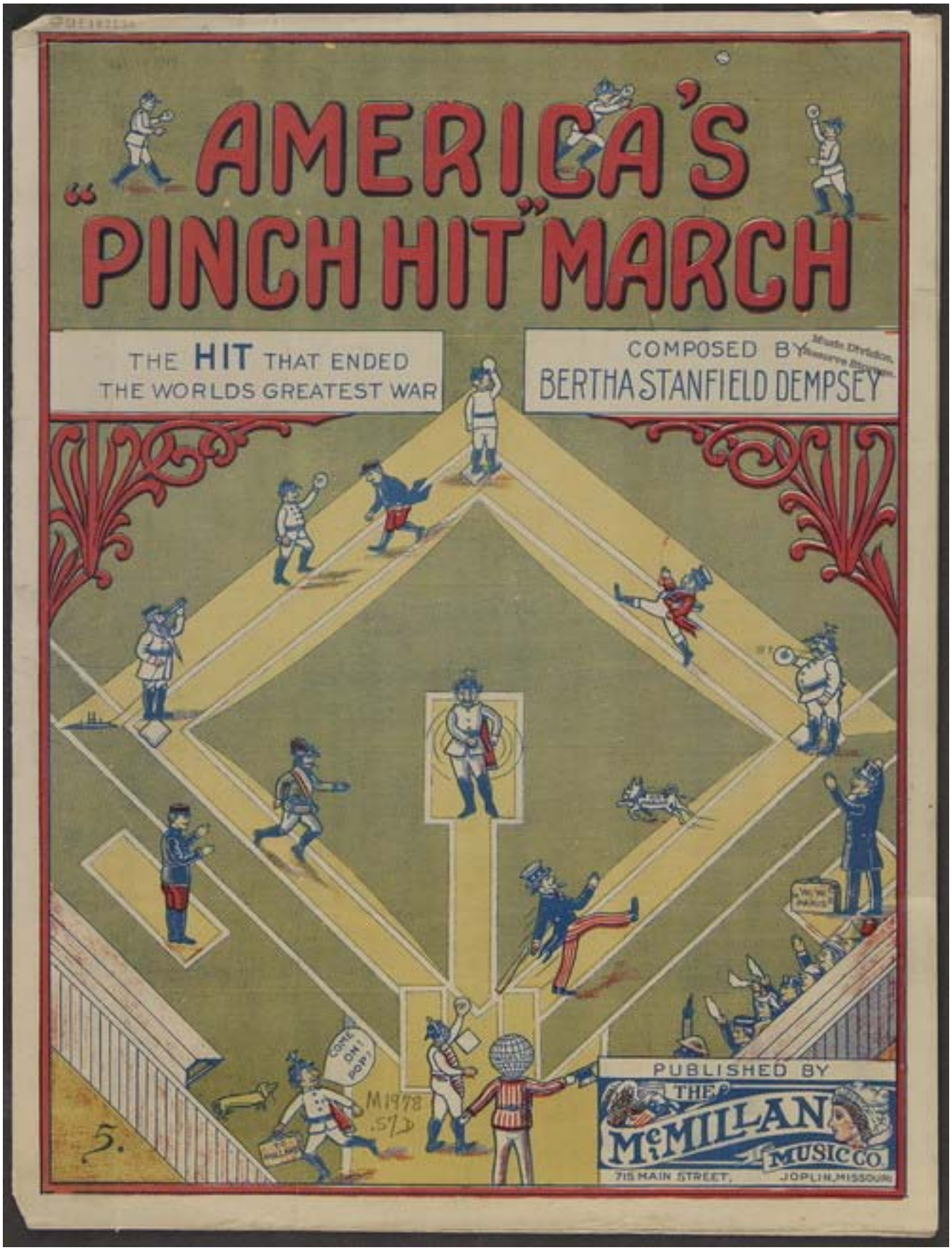
THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Data Element (Field)	Data Value (Content)	Controlled Vocabulary (Yes or No) (Specify, if any)

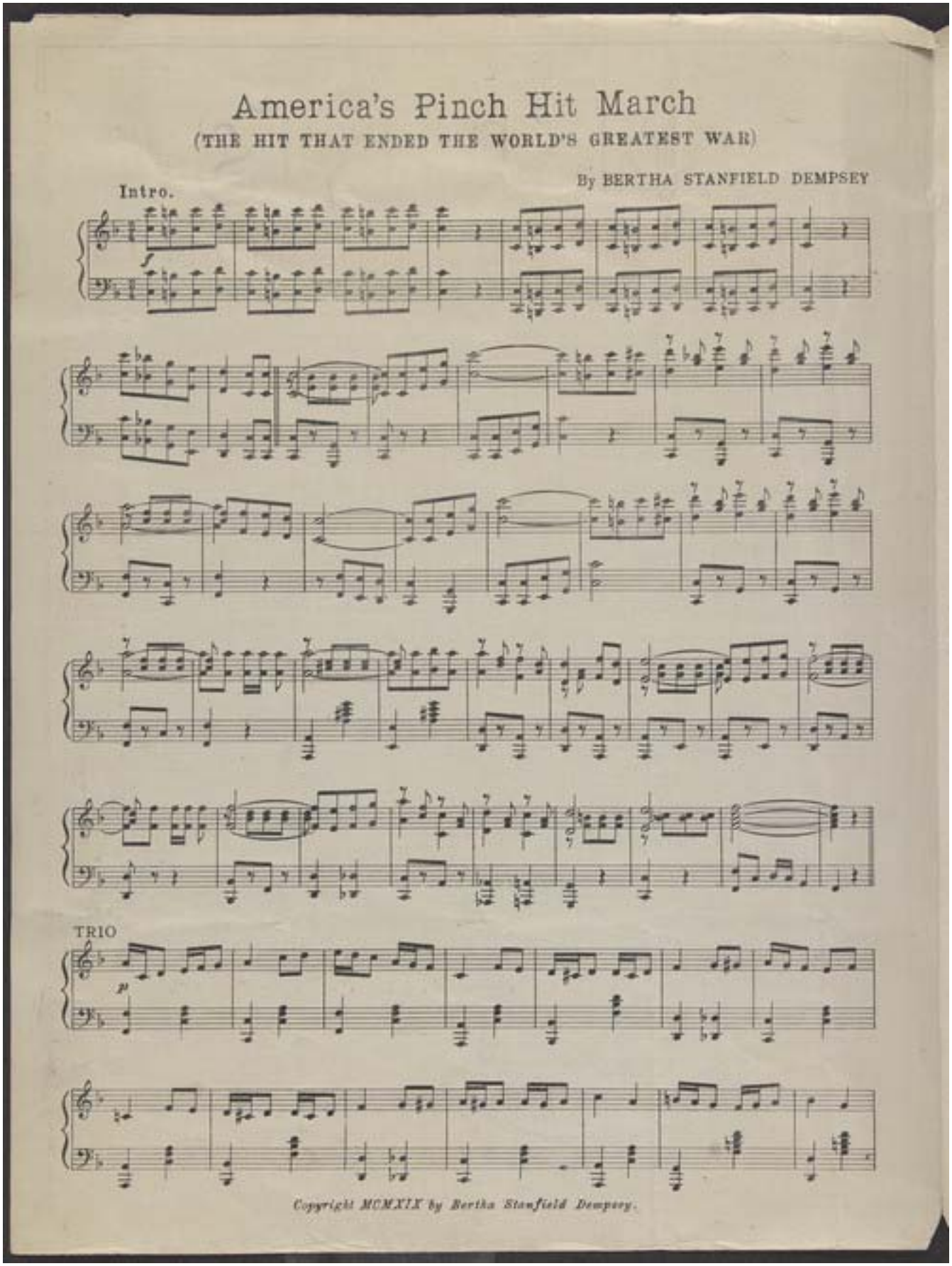
THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Data Element (Field)	Data Value (Content)	Controlled Vocabulary Yes or No Specify, if any

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY


AMERICA'S
"PINCH HIT" MARCH
THE HIT THAT ENDED
THE WORLDS GREATEST WAR
COMPOSED BY
BERTHA STANFIELD DEMPSEY
Music Division
Reserve Section
PUBLISHED BY
THE
McMILLAN
MUSIC CO.
715 MAIN STREET, JOPLIN, MISSOURI
COME ON! POP!
M1978
.S7D
5.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY



America's Pinch Hit March

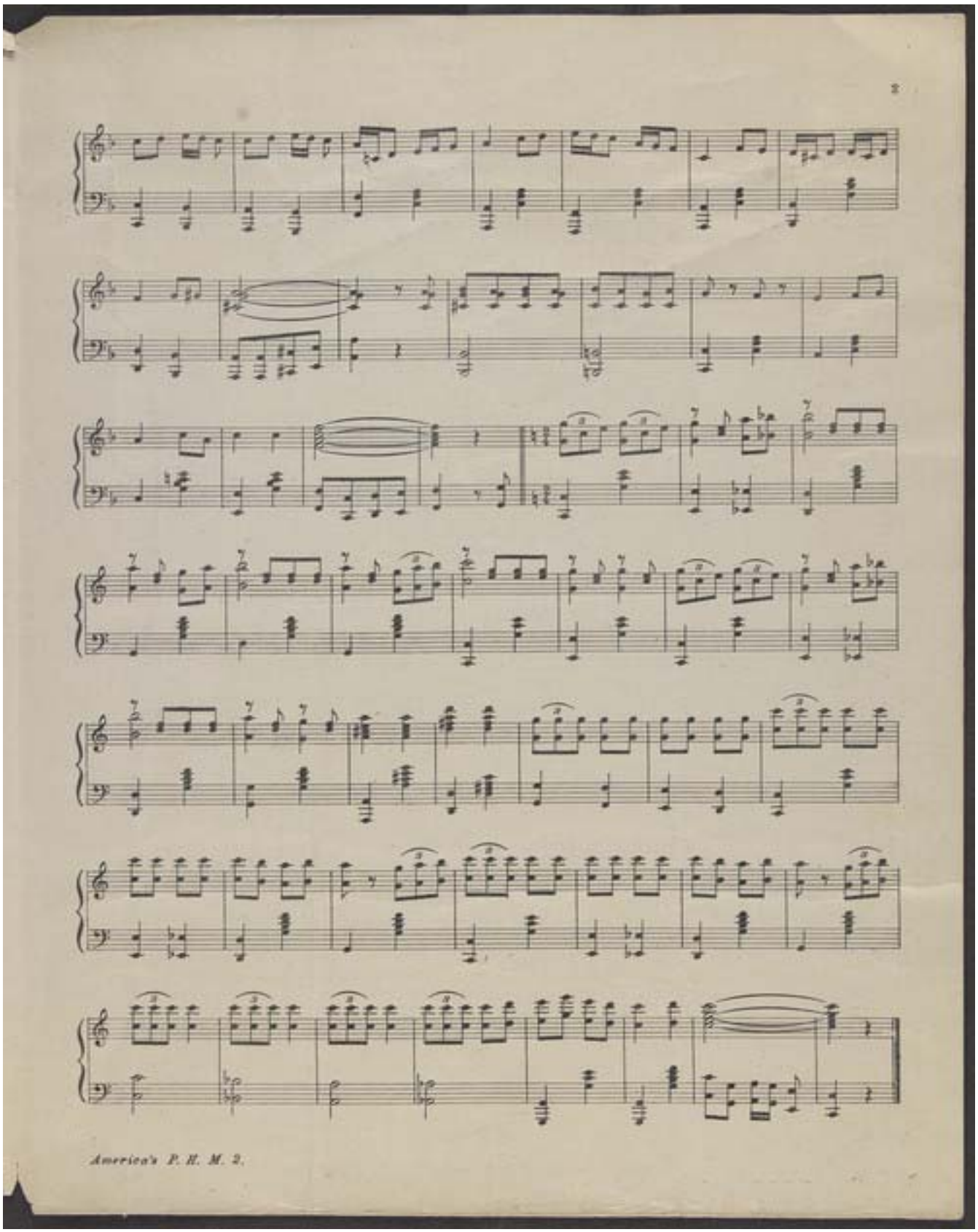
(THE HIT THAT ENDED THE WORLD'S GREATEST WAR)

By BERTHA STANFIELD DEMPSEY

Intro.

TRIO

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY


America's P. H. M. 2.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Dublin Core Metadata Initiative
Making it easier to find information

Dublin Core Metadata Element Set, Version 1.1

Identifier: http://dublincore.org/documents/2008/01/14/dces/

Supersedes: http://dublincore.org/documents/2006/12/18/dces/

Latest version: http://dublincore.org/documents/dces/

Date Issued: 2008-01-14

Status of document: This is a DCMI Recommendation.

Description of document: This document provides ready reference for the Dublin Core Metadata Element Set, Version 1.1. For more detailed documentation and links to historical versioning information, see the document "DCMI Metadata Terms".

Introduction

The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio; "core" because its elements are broad and generic, usable for describing a wide range of resources.

The fifteen element "Dublin Core" described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI). The full set of vocabularies, DCMI Metadata Terms [DCMI-TERMS], also includes sets of resource classes (including the DCMI Type Vocabulary [DCMI-TYPE]), vocabulary encoding schemes, and syntax encoding schemes. The terms in DCMI vocabularies are intended to be used in combination with terms from other, compatible vocabularies in the context of application profiles and on the basis of the DCMI Abstract Model [DCAM].

All changes made to terms of the Dublin Core Metadata Element Set since 2001 have been reviewed by a DCMI Usage Board in the context of a DCMI Namespace Policy [DCMI-NAMESPACE]. The namespace policy describes how DCMI terms are assigned Uniform Resource Identifiers (URIs) and sets limits on the range of editorial changes that may allowably be made to the labels, definitions, and usage comments associated with existing DCMI terms.

This document, an excerpt from the more comprehensive document "DCMI Metadata Terms" [DCTERMS] provides an abbreviated reference version of the fifteen element descriptions that have been formally endorsed in the following standards:

- ISO Standard 15836-2003 of February 2003 [ISO15836]
- NISO Standard Z39.85-2007 of May 2007 [NISOZ3985]
- IETF RFC 5013 of August 2007 [RFC5013]

Since 1998, when these fifteen elements entered into a standardization track, notions of best

practice in the Semantic Web have evolved to include the assignment of formal domains and ranges in addition to definitions in natural language. Domains and ranges specify what kind of described resources and value resources are associated with a given property. Domains and ranges express the meanings implicit in natural-language definitions in an explicit form that is usable for the automatic processing of logical inferences. When a given property is encountered, an inferencing application may use information about the domains and ranges assigned to a property in order to make inferences about the resources described thereby.

Since January 2008, therefore, DCMI includes formal domains and ranges in the definitions of its properties. So as not to affect the conformance of existing implementations of "simple Dublin Core" in RDF, domains and ranges have not been specified for the fifteen properties of the dc: namespace (http://purl.org/dc/elements/1.1/). Rather, fifteen new properties with "names" identical to those of the Dublin Core Metadata Element Set Version 1.1 have been created in the dcterms: namespace (http://purl.org/dc/terms/). These fifteen new properties have been defined as subproperties of the corresponding properties of DCMES Version 1.1 and assigned domains and ranges as specified in the more comprehensive document "DCMI Metadata Terms" [DCTERMS].

Implementers may freely choose to use these fifteen properties either in their legacy dc: variant (e.g., http://purl.org/dc/elements/1.1/creator) or in the dcterms: variant (e.g., http://purl.org/dc/terms/creator) depending on application requirements. The RDF schemas of the DCMI namespaces describe the subproperty relation of dcterms:creator to dc:creator for use by Semantic Web-aware applications. Over time, however, implementers are encouraged to use the semantically more precise dcterms: properties, as they more fully follow emerging notions of best practice for machine-processable metadata.

References

[RFC5013]	http://www.ietf.org/rfc/rfc5013.txt
[NISOZ3985]	http://www.niso.org/standards/resources/Z39-85-2007.pdf
[ISO15836]	http://www.niso.org/international/SC4/n515.pdf
[TRANSLATIONS]	http://dublincore.org/resources/translations/
[DCTERMS]	http://dublincore.org/documents/dcmi-terms/

The Elements

Term Name: contributor	
URI:	http://purl.org/dc/elements/1.1/contributor
Label:	Contributor
Definition:	An entity responsible for making contributions to the resource.
Comment:	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Term Name: coverage	
URI:	http://purl.org/dc/elements/1.1/coverage
Label:	Coverage
Definition:	The spatial or temporal topic of the resource, the spatial applicability of the

	resource, or the jurisdiction under which the resource is relevant.
Comment:	Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN]. Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.
References:	[TGN] http://www.getty.edu/research/tools/vocabulary/tgn/index.html
Term Name: creator	
URI:	http://purl.org/dc/elements/1.1/creator
Label:	Creator
Definition:	An entity primarily responsible for making the resource.
Comment:	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Term Name: date	
URI:	http://purl.org/dc/elements/1.1/date
Label:	Date
Definition:	A point or period of time associated with an event in the lifecycle of the resource.
Comment:	Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].
References:	[W3CDTF] http://www.w3.org/TR/NOTE-datetime
Term Name: description	
URI:	http://purl.org/dc/elements/1.1/description
Label:	Description
Definition:	An account of the resource.
Comment:	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
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/
Term Name: identifier	
URI:	http://purl.org/dc/elements/1.1/identifier
Label:	Identifier
Definition:	An unambiguous reference to the resource within a given context.
Comment:	Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.
Term Name: language	
URI:	http://purl.org/dc/elements/1.1/language
Label:	Language
Definition:	A language of the resource.
Comment:	Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646].
References:	[RFC4646] http://www.ietf.org/rfc/rfc4646.txt
Term Name: publisher	
URI:	http://purl.org/dc/elements/1.1/publisher
Label:	Publisher
Definition:	An entity responsible for making the resource available.
Comment:	Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Term Name: relation	
URI:	http://purl.org/dc/elements/1.1/relation
Label:	Relation
Definition:	A related resource.
Comment:	Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
Term Name: rights	
URI:	http://purl.org/dc/elements/1.1/rights
Label:	Rights

Definition:	Information about rights held in and over the resource.
Comment:	Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.
Term Name: source	
URI:	http://purl.org/dc/elements/1.1/source
Label:	Source
Definition:	A related resource from which the described resource is derived.
Comment:	The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
Term Name: subject	
URI:	http://purl.org/dc/elements/1.1/subject
Label:	Subject
Definition:	The topic of the resource.
Comment:	Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary. To describe the spatial or temporal topic of the resource, use the Coverage element.
Term Name: title	
URI:	http://purl.org/dc/elements/1.1/title
Label:	Title
Definition:	A name given to the resource.
Comment:	Typically, a Title will be a name by which the resource is formally known.
Term Name: type	
URI:	http://purl.org/dc/elements/1.1/type
Label:	Type
Definition:	The nature or genre of the resource.
Comment:	Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.
References:	[DCMITYPE] http://dublincore.org/documents/dcmi-type-vocabulary/

Outline of Elements and Attributes in MODS Version 3.3

This document contains a listing of elements and their related attributes in MODS Version 3.3 with values or value applicable. It is an "outline" of the schema. Items highlighted in red indicate changes made to MODS in Version 3.

All top-level elements and all attributes are optional. Subelements are optional unless indicated as optional. Attribu a mandated sequence and not repeatable (per XML rules). "Ordered" below means the subelements must occur i given. Elements are repeatable unless otherwise noted.

"Authority" attributes are either followed by codes for authority lists (e.g., iso639-2b) or "see" references that link t that contain codes for identifying authority lists.

For additional information about any MODS elements (except for new 3.3 elements), please see the MODS User

Top Level Elements:

Root Elements:

Top Level Elements

1. titleInfo

Subelements:

- title
- subTitle
- partNumber
- partName
- nonSort

Attributes:

ID; xlink; lang; xml:lang; script; transliteration
type (enumerated: abbreviated, translated, alternative, uniform)
authority (see: www.loc.gov/marc/sourcecode/authorityfile/authorityfilesource.html)
displayLabel

2. name

Subelements:

namePart
Attribute: type (date, family, given, termsOfAddress)
displayForm
affiliation
role
roleTerm
Attributes: type (code, text); authority (see: www.loc.gov/marc/sourcecode/relator/relatorsour
description

Attributes:

ID; xlink; lang; xml:lang; script; transliteration
type (enumerated: personal, corporate, conference)
authority (see: www.loc.gov/marc/sourcecode/authorityfile/authorityfilesource.html)

3. typeOfResource

Enumerated values:

text
cartographic
notated music
sound recording-musical
sound recording-nonmusical
sound recording
still image
moving image
three dimensional object
software, multimedia
mixed material

Subelements:

[none]

Attributes:

collection (yes)
manuscript (yes)

4. genre

Subelements:

[none]

Attributes:

lang; xml:lang; script; transliteration
authority (see: www.loc.gov/marc/sourcecode/genre/genresource.html)
type (examples: class, work type, or style)

5. originInfo

Subelement:

place

placeTerm

Attributes: type (code, text); authority (marcgac, marccountry, iso3166)

publisher

dateIssued

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxim questionable)

dateCreated

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxim questionable)

dateCaptured

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxim questionable)

dateValid

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxim questionable)

dateModified

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxim questionable)

copyrightDate

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxim questionable)

dateOther

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxim questionable); type

edition

issuance (continuing, monographic)

frequency

Attribute: authority (see: www.loc.gov/marc/sourcecode/frequency/frequencyhome.html)

Attributes:

lang; xml:lang; script; transliteration

6. language

Subelements:

languageTerm

Attributes: type (code, text); authority (iso639-2b, rfc3066, iso639-3, rfc4646)

Attributes:

objectPart

7. physicalDescription

Subelements:

form

Attribute: authority (see: www.loc.gov/marc/sourcecode/form/formsource.html); type (Examples technique)

reformattingQuality (access, preservation, replacement)

internetMediaType

extent

digitalOrigin (born digital, reformatted digital, digitized microfilm, digitized other analog)
note
Attributes: xlink; lang; xml:lang; script; transliteration; displayLabel; type (For a list of implement see: www.loc.gov/standards/mods/mods-notes.html)

Attributes:
lang
xml:lang
script
transliteration

8. abstract

Subelements:
[none]

Attributes:
xlink; lang; xml:lang; script; transliteration
displayLabel
type (Examples: review, scope and content)

9. tableOfContents

Subelements:
[none]

Attributes:
xlink; lang; xml:lang; script; transliteration
displayLabel
type (Examples: incomplete contents, partial contents)

10. targetAudience

Subelements:
[none]

Attributes:
lang; xml:lang; script; transliteration
authority (see: www.loc.gov/marc/sourcecode/target/targetsource.html)

11. note

Subelements:
[none]

Attributes:
ID; xlink; lang; xml:lang; script; transliteration
displayLabel
type (For a list of implemented note types, see: www.loc.gov/standards/mods/mods-notes.html)

12. subject

Subelements:
topic
geographic
temporal
Attributes: encoding (w3cdtf, iso8601, marc); point (start,end); keyDate (yes); qualifier (approxir questionable)

titleInfo (see: titleInfo)
name (see: name)
geographicCode
Attribute: authority (marcgac, marccountry, iso3166)
genre
hierarchicalGeographic
continent
country
province
region
state
territory
county
city
island
area
extraterrestrialArea
citySection
cartographics [ordered]
scale
projection
coordinates
occupation

Attributes:

ID; xlink; lang; xml:lang; script; transliteration
authority (see: www.loc.gov/marc/sourcecode/subject/subjectsource.html)

13. classification

Subelements:

[none]

Attributes:

lang; xml:lang; script; transliteration
authority (see: www.loc.gov/marc/sourcecode/classification/classificationsource.html)
edition
displayLabel

14. relatedItem

Subelements:

(Any MODS element may be used as defined in the schema with appropriate subelements.)
titleInfo
name
typeOfResource
genre
originInfo
language
physicalDescription
abstract

tableOfContents
targetAudience
note
subject
classification
relatedItem
identifier
location
accessCondition
part
extension
recordInfo

Attributes:

ID; xlink

displayLabel

type (enumerated: preceding, succeeding, original, host, constituent, series, otherVersion, otherFo
isReferencedBy)

15. identifier

Subelements:

[none]

Attributes:

lang; xml:lang; script; transliteration

type
(suggested values: hdl, doi, isbn, isrc, ismn, issn, issue number, istc, lccn, local, matrix number, mu
music plate, sici, uri, upc, videorecording identifier, stock number)

invalid (yes)

16. location

Subelements:

physicalLocation

Attributes: authority (see: www.loc.gov/marc/sourcecode/organization/organizationsource.html
displayLabel; type (Examples: current, discovery, former, creation); lang; xml:lang; script; trans

shelfLocator

url

Attributes:

dateLastAccessed
displayLabel
note
access (preview, raw object, object in context)
usage (primary display)

holdingSimple (not repeatable)

copyInformation

form (not repeatable)

Attribute: authority

sublocation

shelfLocator

electronicLocator
note
Attributes: displayLabel, type
enumerationAndChronology
Attributes: unitType (1,2,3)
Note: 1=basic bibliographic unit; 2=supplement; 3=index

holdingExternal (not repeatable)
(Extensible to use other holdings schemas)

17. accessCondition

(Extensible to allow for other more detailed rights schemas.)
Subelements:
[none]
Attributes:
xlink; lang; xml:lang; script; transliteration
displayLabel
type (suggested values: restriction on access; use and reproduction)

18. part

Subelements:
detail
number
caption
title
Attributes: type (suggested values: part, volume, issue, chapter, section, paragraph, track) ; level
extent [ordered]
start
end
total
list
Attribute: unit (suggested values: pages, minutes)
date
Attributes: encoding (w3cdtf, iso8601, marc); point (start,end); qualifier (approximate, inferred, ques
text
Attributes: xlink; lang; xml:lang; transliteration; script; displayLabel; type

Attributes:
ID
type (suggested values: volume, issue, chapter, section, paragraph, track)
order

19. extension

Subelements:
[none]
Attributes:
[none]

20. recordInfo

Subelements:

recordContentSource

Attributes: authority (see: www.loc.gov/marc/sourcecode/organization/organizationsource.html); la
script; transliteration

recordCreationDate

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxima
questionable)

recordChangeDate

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approxima
questionable)

recordIdentifier

Attribute: source

recordOrigin

languageOf Cataloging

languageTerm

Attributes: type (code, text); authority (iso639-2b, rfc3066)

descriptionStandard (see: http://www.loc.gov/marc/relators/reladesc.html#rela040b)

Attributes:

lang; xml:lang; script; transliteration

Root Elements

1. mods (A single MODS record)

Subelements:

See: Top Level Elements

Attributes:

ID

version

2. modsCollection (A collection of MODS records)

Subelements:

mods

Attributes:

[none]

HOME >> Schemas >> **Outline of Elements and Attributes**

Questions and comments:
Contact Us (January 24, 2008)

VRA Core 4.0 Outline

Global Attributes

- dataDate
- extent
- href
- pref
- refid
- rules
- source
- vocab
- xml:lang

Syntax (using date element as example)

```
<work id="">
    <dateSet>
        <display></display>
        <notes></notes>
        <date type="">
            <earliestDate></earliestDate>
            <latestDate></latestDate>
        </date>
    </dateSet>
</work>
```

ELEMENTS

- **work, collection, or image** *(id)*
- **agent**
  - attribution
  - culture
  - dates (*type*)
    - earliestDate *(circa)*
    - latestDate *(circa)*
  - name (*type*)
  - role
- **culturalContext**
- **date** (*type*)
  - earliestDate *(circa)*
  - latestDate *(circa)*
- **description**
- **inscription**
  - author
  - position
  - text (*type*)
- **location** *(type)*
  - name (*type*)
  - refid (*type*)
- **material** (*type*)
- **measurements** (*type, unit*)
- **relation** (*type, relids*)
- **rights** *(type)*
  - rightsHolder
  - text
- **source**
  - name (*type*)
  - refid (*type*)
- **stateEdition** (*count, num, type*)
  - description
  - name
- **stylePeriod**
- **subject**
  - term *(type)*
- **technique**
- **textref**
  - name (*type*)
  - refid *(type)*
- **title** (*type*)
- **worktype**

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

PERFORMING ARTS ENCYCLOPEDIA

Katherine Dunham at the Theâtre de Paris in 1948 / Raymond Voi [photograph]

Title
Katherine Dunham at the Theâtre de Paris in 1948 [photograph]

Photographer
Raymond Voinquel

Form
photograph

Permissions note
Special Collections Research Center, Morris Library, Southern Illinois University, Carbondale

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

3. Technical and administrative metadata standards

Metadata Standards and Applications Workshop

1

Goals of session

- To understand the different types of administrative metadata standards
- To learn what types of metadata are needed for digital preservation
- To learn the importance of technical, structural and rights metadata in digital libraries

2

Types of administrative metadata

- Provides information to help manage a resource
  - Preservation metadata
    - Technical characteristics
    - Information about actions on an object
  - Structural metadata may be considered administrative; indicates how compound objects are put together
  - Rights metadata
    - Access rights and restrictions
    - Preservation rights and restrictions

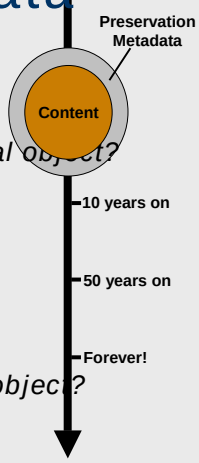
3

PREMIS: introduction

- Preservation metadata that includes subcategories:
  - Technical metadata
  - Relationships (structural and derivative)
  - Digital provenance (what actions performed on objects
  - Rights

4

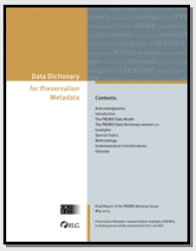
Preservation metadata includes:



- Provenance:
  - *Who has had custody/ownership of the digital object?*
- Authenticity:
  - *Is the digital object what it purports to be?*
- Preservation Activity:
  - *What has been done to preserve the digital object?*
- Technical Environment:
  - *What is needed to render and use the digital object?*
- Rights Management:
  - *What IPR must be observed?*

5

PREMIS Data Dictionary



- **May 2005:**
  Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group
- 237-page report includes:
  - PREMIS Data Dictionary 1.0
  - Accompanying report
  - Special topics, glossary, usage examples
- **Data Dictionary:** comprehensive, practical resource for implementing preservation metadata in digital archiving systems
  - Used *Framework* as starting point
  - Detailed description of metadata elements
  - Guidelines to support implementation, use, management
  - Based on deep pool of institutional experiences in setting up and managing operational capacity for digital preservation
- Set of **XML schema** developed to support use of Data Dictionary

Scope of data dictionary

- Implementation independent
- Descriptive metadata out of scope
- Technical metadata applying to all or most format types
- Media or hardware details are limited
- Business rules are essential for working repositories, but not covered
- Rights information for preservation actions, not access

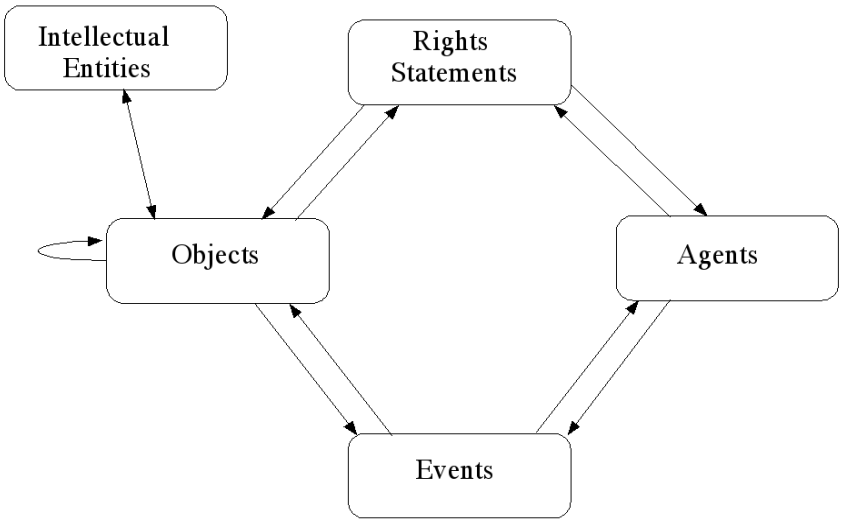
7

What PREMIS is and is not

- What PREMIS is:
  - Common data model for organizing/thinking about preservation metadata
  - Guidance for local implementations
  - Standard for exchanging information packages between repositories
- What PREMIS is not:
  - Out-of-the-box solution: need to instantiate as metadata elements in repository system
  - All needed metadata: excludes business rules, format-specific technical metadata, descriptive metadata for access, non-core preservation metadata
  - Lifecycle management of objects outside repository
  - Rights management: limited to permissions regarding actions taken within repository

8

PREMIS Data Model



9

Types of information covered in PREMIS (by entity type)

- Object
  - Object ID
  - Preservation level
  - Object characteristics (format, size, etc.)
  - Storage
  - Environment
  - Digital signatures
  - Relationships
  - Linking identifiers
- Event
  - Event ID
  - Event type
  - Event date/time
  - Event outcomes
  - Linking identifiers
- Agent
  - Agent ID
  - Agent name
- Rights
  - Rights statement
  - Granting agent
  - Permission granted

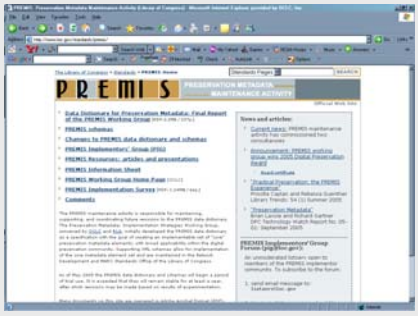
10

PREMIS Maintenance Activity

Permanent Web presence, hosted by Library of Congress

Centralized destination for information, announcements, and other PREMIS-related resources

Discussion list for PREMIS implementers (PIG list)



Coordinate future revisions of Data Dictionary and XML schema

Editorial committee guides development and revisions

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

11

Current activities

- PREMIS Implementers' Registry
  - http://www.loc.gov/standards/premis/premis-registry.html
- Revision of data dictionary and schemas (March 2008)
- Guidelines for use of PREMIS within METS
- PREMIS tutorials
  - One or one and a half day tutorials have been given in several locations: Glasgow, Boston, Stockholm, Albuquerque, Washington, San Diego
  - Training materials available from LC

12

Why is PREMIS important to catalogers?

- As we take responsibility for more digital materials, we need to ensure that they can be used in the future
- Most preservation metadata will be generated from the object, but catalogers may need to verify its accuracy
- Catalogers may need to play a role in assessing and organizing digital materials
  - Understanding the structure of complex digital objects
  - Determining significant properties that need to be preserved

13

Technical metadata for images

- NISO Z39.87 and MIX
- Adobe and XMP
- Exif
- IPTC (International Press Telecommunications Council)/XMP
- Some of these deal with embedded metadata in images

14

Metadata For Images in XML (MIX)

- An XML Schema designed for expressing technical metadata for digital still images
- Based on the NISO Z39.87 Data Dictionary – Technical Metadata for Digital Still Images
- Can be used standalone or as an extension schema with METS/PREMIS

15

Using MIX

- Technical metadata for digital images
  - Characteristics that apply to all or most object types, e.g. size, format (elements also in PREMIS)
  - Format specific metadata for images, e.g. bit depth, color space, etc.
- Originally designed for TIFF files
- Recent revision (version 1.0) includes support for GIS images and JPEG 2000; data element names harmonized with those in PREMIS
- Most well developed of format specific technical metadata standards
- Example of MIX metadata

16

Technical metadata for textual objects

- textMD is an XML Schema designed for expressing technical metadata for textual objects
- Developed at New York University; maintenance transferred to LC
- Includes format specific technical metadata for text, e.g. byte order, character set encoding, font script

17

Technical metadata for audio and video

- Not as well developed as other technical metadata
- Complexities of file formats requires expertise to develop these
- LC developed XML technical metadata schemas in 2003/2004 for LC Audiovisual Prototype Project used with METS; these were widely implemented because of the lack of other schemas
- Audio and video technical metadata schemas under development by expert organizations

18

Technical metadata for audio

- Audio Engineering Society
- Audio Objects (AES-098B) schema
- Accommodates analog and digital as well as segmenting
- Schema for process history metadata
- Integrates audio and video to some extent

19

Technical metadata for moving image

- Society of Motion Picture and Television Engineers (SMPTE) standards: technical standards for television, motion picture and digital cinema
- MPEG family of standards: MPEG-7
- These are highly technical and may be difficult to understand and use
- Moving Image Catalog (MIC) project is also experimenting with these

20

Structural metadata

- Supports the intended presentation and use and navigation of an object
- Binds the parts together; expresses relationships between parts of a multipart object
- Examples of structural metadata expressions:
  - METS structMap
  - PREMIS relationship elements
  - EAD hierarchical structure

21

Rights metadata

- Rights schemas with limited scope
- Full blown Rights Expression Languages (REL) for managing intellectual property rights, particularly by rights owners
- Rights information is not well understood
  - Different laws in different jurisdictions
  - Machine actionable vs. human understandable
- Rights take different forms
  - legal statutes, e.g. copyright
  - contractual rights, e.g. licenses

22

Rights schemas with limited scope

- METS Rights
  - Access rights for use with METS objects
  - Rights declarations
  - Rights holder
  - Context
- CDL copyright schema
  - Specifically copyrights, not other intellectual property rights
  - Information you need to know to assess copyright status (e.g. creators, rights holders, dates, jurisdiction)

23

Rights schemas with limited scope cont.

- PREMIS Rights
  - Focused on rights for preservation rather than access
  - Revision of PREMIS data dictionary expanded this area
  - Allows for extensibility, i.e. inserting another rights schema
- Creative commons
  - Allows creators to choose a license for their work
  - Simple rights statements that fit a lot of situations

24

Rights metadata for specific object types

- PLUS for images
- MPEG-21 REL for moving images, etc.
- ONIX for licensing terms
- Full Rights Expression Languages
  - XRML/MPEG 21
  - ODRL

25

Exercise

- Provide administrative/technical metadata for the object used in the descriptive metadata exercise

26

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Session 3: Administrative metadata exercise (PREMIS)
Information needed to fill in template

Example 1 (sheet music):

America's pinch hit march
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200033287/default.html

Sheet music; 3 pages (2 pages music with cover)
File 1 (cover):
Filename: FN10057
Full path:
http://lcweb2.loc.gov/natlib/ihas/service/encyclopedia/200033287/0001v.jpg
File format: image/jpeg
Size: 629507
Software used to access: Macromedia Fireworks MX Version 6.0
File created: 2 January 2008
Application used for creation: Adobe photoshop version CS3

File 2 (1st page of music):
Filename: FN10075
Full path:
http://lcweb2.loc.gov/natlib/ihas/service/encyclopedia/200033287/0002v.jpg
File format: image/jpeg
Size: 399565
Software used to access: Macromedia Fireworks MX Version 6.0
File created: 2 January 2008
Application used for creation: Adobe photoshop version CS3

Example 2 (photograph):

Photograph of Katherine Dunham at the Theâtre de Paris in 1948.
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200003731/default.html

File 1 (master TIFF)
Filename: FN10029
Full path:
http://lcweb2.loc.gov/natlib/ihas/warehouse/dunham/200003731/ver01/0001.tif
File format: image/tiff
Size: 9633805
Software used to access: Firefox version 5.0
File created: May 29, 1999
Application used for creation: Adobe photoshop version CS3

File 2 (derivative JPEG)
Filename: FN1005C
Full path:

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

http://lcweb2.loc.gov/natlib/ihas/service/dunham/200003731/ver01/0001v.jpg
File format: image/jpeg
Size: 270400
Software used to access: Firefox version 5.0
File created: May 29, 1999
Application used for creation: Adobe photoshop version CS3

Use metadata template provided.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Session 3: Administrative metadata exercise (PREMIS)
Information needed to fill in template

Example 1 (sheet music):

America's pinch hit march
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200033287/default.html

Sheet music; 3 pages (2 pages music with cover)
File 1 (cover):
Filename: FN10057
Full path:
http://lcweb2.loc.gov/natlib/ihas/service/encyclopedia/200033287/0001v.jpg
File format: image/jpeg
Size: 629507
Software used to access: Macromedia Fireworks MX Version 6.0
File created: 2 January 2008
Application used for creation: Adobe photoshop version CS3

File 2 (1st page of music):
Filename: FN10075
Full path:
http://lcweb2.loc.gov/natlib/ihas/service/encyclopedia/200033287/0002v.jpg
File format: image/jpeg
Size: 399565
Software used to access: Macromedia Fireworks MX Version 6.0
File created: 2 January 2008
Application used for creation: Adobe photoshop version CS3

Example 2 (photograph):

Photograph of Katherine Dunham at the Theâtre de Paris in 1948.
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200003731/default.html

File 1 (master TIFF)
Filename: FN10029
Full path:
http://lcweb2.loc.gov/natlib/ihas/warehouse/dunham/200003731/ver01/0001.tif
File format: image/tiff
Size: 9633805
Software used to access: Firefox version 5.0
File created: May 29, 1999
Application used for creation: Adobe photoshop version CS3

File 2 (derivative JPEG)
Filename: FN1005C
Full path:

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

http://lcweb2.loc.gov/natlib/ihas/service/dunham/200003731/ver01/0001v.jpg
File format: image/jpeg
Size: 270400
Software used to access: Firefox version 5.0
File created: May 29, 1999
Application used for creation: Adobe photoshop version CS3

Use metadata template provided.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Selected PREMIS elements

- objectIdentifier
  - objectIdentifierType
  - objectIdentifierValue
- objectCharacteristics
  - size
  - format
    - formatDesignation
      - formatName
      - formatVersion
- significantProperties
- creatingApplication
  - creatingApplicationName
  - creatingApplicationVersion
  - dateCreatedByApplication
-
- environment
  - software
    - swName
    - swVersion
    - swType
- relationship
  - relationshipType
  - relationshipSubType
  - relatedObjectIdentification
    - relatedObjectIdentifierType
    - relatedObjectIdentifierValue
    - relatedObjectSequence

Controlled vocabularies:

RelationshipType:
structural = a relationship between parts of an object

derivation = a relationship where one object is the result of a transformation performed on the related object

RelationshipSubType:
is child of = the object is directly subordinate in a hierarchy to the related object (Note that this is semantically equivalent to “Has parent,” which may be preferred by some implementations.

is parent of = the object is directly superior in a hierarchy to the related object (Note that this is semantically equivalent to “Has child,” which may be preferred by some implementations.

Selected PREMIS elements

- objectIdentifier
  - objectIdentifierType
  - objectIdentifierValue
- objectCharacteristics
  - size
  - format
    - formatDesignation
      - formatName
      - formatVersion
- significantProperties
- creatingApplication
  - creatingApplicationName
  - creatingApplicationVersion
  - dateCreatedByApplication
-
- environment
  - software
    - swName
    - swVersion
    - swType
- relationship
  - relationshipType
  - relationshipSubType
  - relatedObjectIdentification
    - relatedObjectIdentifierType
    - relatedObjectIdentifierValue
    - relatedObjectSequence

Controlled vocabularies:

RelationshipType:
structural = a relationship between parts of an object

derivation = a relationship where one object is the result of a transformation performed on the related object

RelationshipSubType:
is child of = the object is directly subordinate in a hierarchy to the related object (Note that this is semantically equivalent to “Has parent,” which may be preferred by some implementations.

is parent of = the object is directly superior in a hierarchy to the related object (Note that this is semantically equivalent to “Has child,” which may be preferred by some implementations.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

has sibling = the object shares a common parent with the related object

is part of = the object is contained by the related object

has part = the object contains the related object

source of = the related object is a version of this object created by a transformation

has root = for a representation only, the related object is the file that must be processed first in order to render the representation

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

4. Metadata syntaxes and containers

Metadata Standards and Applications Workshop

1

Goals of session

- Understand syntaxes used for encoding information, including HTML, XML and RDF
- Discover how container formats are used for managing digital resources and their metadata

2

Overview of Syntaxes

- HTML, XHTML: Hypertext Markup Language; eXtensible Hypertext Markup Language
- XML: Extensible Markup Language
- RDF/XML: Resource Description Framework

3

HTML

- HyperText Markup Language
- HTML 4 is the current standard
- HTML is an SGML (Standard Generalized Markup Language) application conforming to International Standard ISO 8879
- Widely regarded as the standard publishing language of the World Wide Web
- HTML addressed the problem of SGML complexity by specifying a small set of structural and semantic tags suitable for authoring relatively simple documents

4

XHTML

- XML-ized version of HTML 4.0, tightens up HTML to match XML syntax
  - Requires ending tags, quoted attributes, lower case, etc., to conform to XML requirements
- XHTML is a W3C specification, redefining HTML as an XML implementation, rather than an SGML implementation
- Imposes requirements that are intended to lead to more well-formed, valid XML, easier for browsers to handle

5

An XHTML Example

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.title" content="Using Dublin Core" />
<meta name="DC.creator" content="Diane Hillmann" />
<meta name="DC.subject" content="documents; Bibliography; Model; meta; Glossary; mark; matching;
refinements; XHTML; Controlled; Qualifiers; Hillmann; mixing; encoding; Diane; Issues; Appendix; elements;
Simple; Special; element; trademark/service; DCMI; Dublin; pages; Section; Resource; Grammatical; Qualified;
XML; Using; Principles; Documents; licensing; OCLC; formal; Usageguide; Roles; Implementing; Contents;
Guidelines; Expressing; Table; Syntax; Content; Element; DC.dot; Home; document; Metadata; RDF/XML;
Website; metadata; privacy; schemes; liability; profiles; Elements; Copyright; Localization; schemas;
HTML/XHTML; Core; Guide; registry; Research; contact; Scope; Projects; languages; Maintenance; Application;
available; Internationalization; HTML; Recommended; link; Purpose; Abstract; AskDCMI; Vocabularies; software;
Storage; Introduction" />
<meta name="DC.description" content="This document is intended as an entry point for users of Dublin Core. For
non-specialists, it will assist them in creating simple descriptive records for information resources (for example,
electronic documents). Specialists may find the document a useful point of reference to the documentation of
Dublin Core, as it changes and grows." />
<meta name="DC.publisher" content="Dublin Core Metadata Initiative" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
<meta name="DC.format" content="text/html" />
<meta name="DC.format" content="31250 bytes" />
<meta name="DC.identifier" scheme="DCTERMS.URI" content="http://dublincore.org/documents/usageguide/" />
```

6

XML

- Extensible Markup Language
- A '*metamarkup' language*: has no fixed tags or elements
- Strict grammar imposes structure designed to be read by machines
- Two levels of conformance:
  - well-formed--conforms to general grammar rules
  - valid--conforms to particular XML schema or DTD (document type definition)

7

XML: Extensible Markup Language

- A technical approach to convey meaning with data
- Not a natural language, although uses natural languages
  - <姓名>Louis Armstrong</姓名>
  - <name>Louis Armstrong</name>
- Not a programming language
- Language in the sense of:
  - A limited set of tags defines the elements that can be used to markup data
  - The set of tags and their relationships need to be explicitly defined (e.g., in XML schema)
  - We can build software that uses XML as input and processes them in a meaningful way
- You can define your own markups and schemas

8

XML is the *lingua franca* of the Web

- Web pages increasingly use at least XHTML
- Business use for data exchange/ messaging
- Family of technologies can be leveraged
  - XML Schema, XSLT, XPath, and XQuery
- Software tools widely available (open source)
  - Storage, editing, parsing, validating, transforming and publishing XML
- *Microsoft Office 2003* supports XML as document format (WordML and ExcelML)
- *Web 2.0* applications based on XML (AJAX, Semantic Web, Web Services, etc.)

9

An XML Schema Defines:

- What elements may be used
- Of which types
- Any attributes
- In which order
- Optional or compulsory
- Repeatability
- Subelements
- Logic

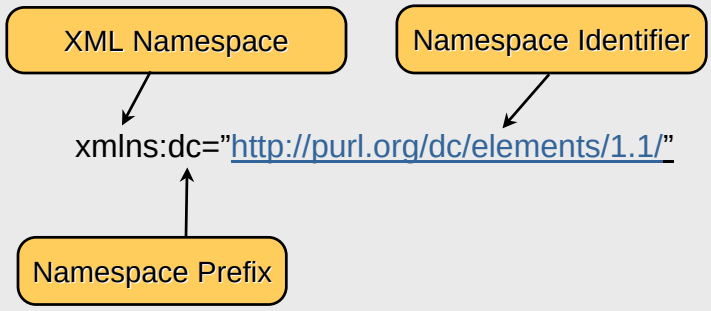
10

Anatomy of an XML Record

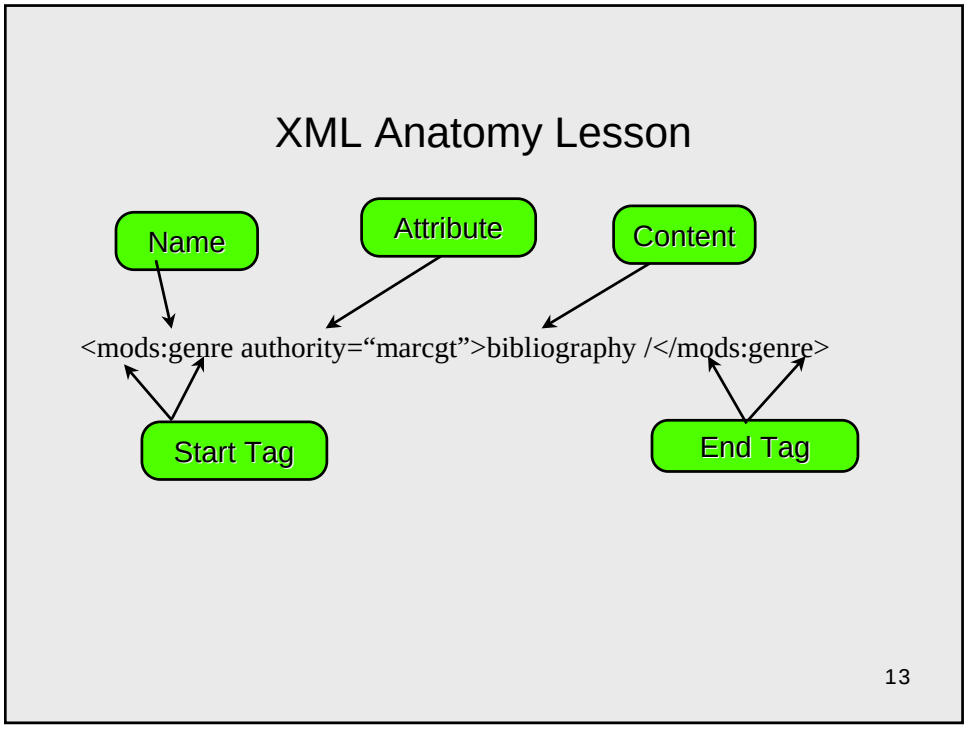
- XML declaration--prepares the processor to work with the document and states the XML version
- Namespaces (uses xmlns:*prefix* and a URI to attach a prefix to each element and attribute)
  - Distinguishes between elements and attributes from different vocabularies that might share a name (but not necessarily a definition) using association with URIs
  - Groups all related elements from an application so software can deal with them
  - The URIs are the standardized bit, not the prefix, and they don't necessarily lead anywhere useful, even if they look like URLs

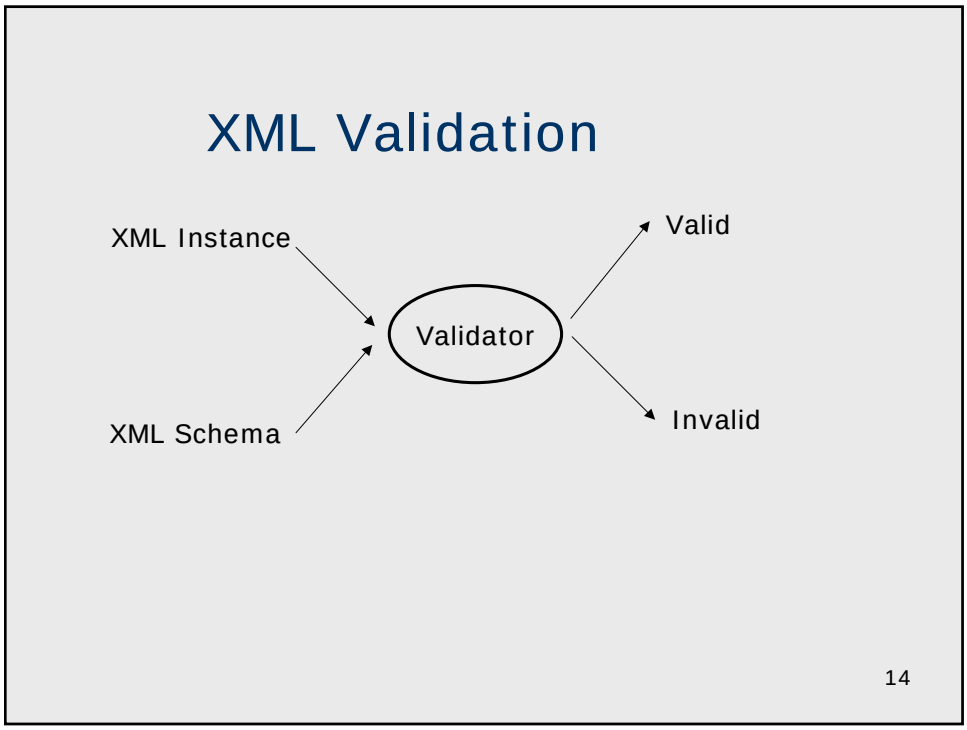
11

XML Namespaces



12


XML Anatomy Lesson
Name
Attribute
Content
<mods:genre authority="marcgt">bibliography /</mods:genre>
Start Tag
End Tag
13


XML Validation
XML Instance
XML Schema
Validator
Valid
Invalid
14

XML Schema Example

```
<xs:element name="software" minOccurs="0"
   maxOccurs="unbounded">
   <xs:complexType>
        <xs:sequence>
             <xs:element name="swName" minOccurs="1"
   maxOccurs="1" type="xs:string"></xs:element>
             <xs:element name="swVersion" minOccurs="0"
   maxOccurs="1" type="xs:string"></xs:element>
             <xs:element name="swType" minOccurs="1"
   maxOccurs="1" type="xs:string"></xs:element>
             <xs:element name="swOtherInformation"
   minOccurs="0" maxOccurs="unbounded" type="xs:string">
   </xs:element>
             <xs:element name="swDependency" minOccurs="0"
   maxOccurs="unbounded" type="xs:string"> </xs:element>
        </xs:sequence>
   </xs:complexType>
</xs:element>
```

15

Will the following XML instance validate?

```
<software>
   <swName>Windows</swName>
   <swVersion>2000</swVersion>
   <swType>Operating System</swType>
</software>
```

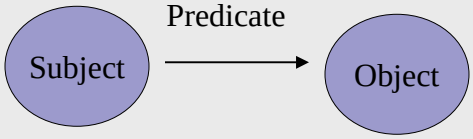
How about this?

```
<swVersion>2000</swVersion>
```

16

Resource Description Framework

- Application for describing and exchanging metadata
- Structure consists of “triples”
- Designed to be read by computers, not humans
- An ontology language to support semantic interoperability—understanding meanings
- Considered an essential part of the Semantic Web
- Can be expressed using XML



http://www.w3.org/RDF

17

Some RDF Concepts

- A **Resource** is anything you want to describe
- A **Class** is a category; it is a set that comprises individuals
- A **Property** is a Resource that has a name, such as "creator" or "homepage"
- A **Property value** is the value of a Property, such as “Barack Obama" or "http://dublincore.org" (note that a property value can be another resource)

18

RDF Statements

- The combination of a **Resource**, a **Property**, and a **Property value** forms a *Statement* (known also as the subject, predicate and object of a Statement), also known as “triples”
- An example *Statement*: "The editor of http://dublincore.org/documents/usageguide/ is Diane Hillmann"
- The subject of the statement above is: http://dublincore.org/documents/usageguide/
- The predicate is: editor
- The object is: Diane Hillmann

19

RDF and OWL

- RDF does not have the language to specify all relationships
- Web Ontology Language (OWL) can specify richer relationships, such as equivalence, inverse, unique
- RDF and OWL may be used together
- RDFS: a syntax for expressing relationships between elements

20

An RDF/XML Example (with DC)

```
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dc="http://purl.org/dc/elements/1.1/">
   <rdf:Description rdf:about="http://www.dlib.org">
    <dc:title>D-Lib Program - Research in Digital Libraries</dc:title>
    <dc:description>The D-Lib program supports the community of people
     with research interests in digital libraries and electronic
     publishing.</dc:description>
    <dc:publisher>Corporation For National Research Initiatives</dc:publ
    <dc:date>1995-01-07</dc:date>
    <dc:subject>
      <rdf:Bag>
        <rdf:li>Research; statistical methods</rdf:li>
        <rdf:li>Education, research, related topics</rdf:li>
        <rdf:li>Library use Studies</rdf:li>
      </rdf:Bag>
    </dc:subject>
    <dc:type>World Wide Web Home Page</dc:type>
    <dc:format>text/html</dc:format>
    <dc:language>en</dc:language>
   </rdf:Description>
</rdf:RDF>
```

Note unordered list

21

Overview of container formats

- A container format is needed to package together all forms of metadata and digital content
- Use of a container is compatible with and an implementation of the OAIS information package concept
- METS: packages metadata with objects or links to objects and defines structural relationships
- MPEG 21 DID: represents digital objects using a flexible and expressive model

22

Metadata Encoding & Transmission Standard

- Developed by the Digital Library Federation, maintained by the Library of Congress
- "... an XML document format for encoding metadata necessary for both management of digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users)."
- Records the (possibly hierarchical) structure of digital objects, the names and locations of the files that comprise those objects, and the associated metadata

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

23

What is METS used for?

- To package metadata with digital object in XML syntax
- For retrieving, storing, preserving, serving resource
- For interchange of digital objects with metadata
- As information package in a digital repository (may be a unit of storage or a transmission format)

24

Characteristics of METS

- Open non-proprietary standard
- Extensible
- Modular
- Developed by the digital library community

25

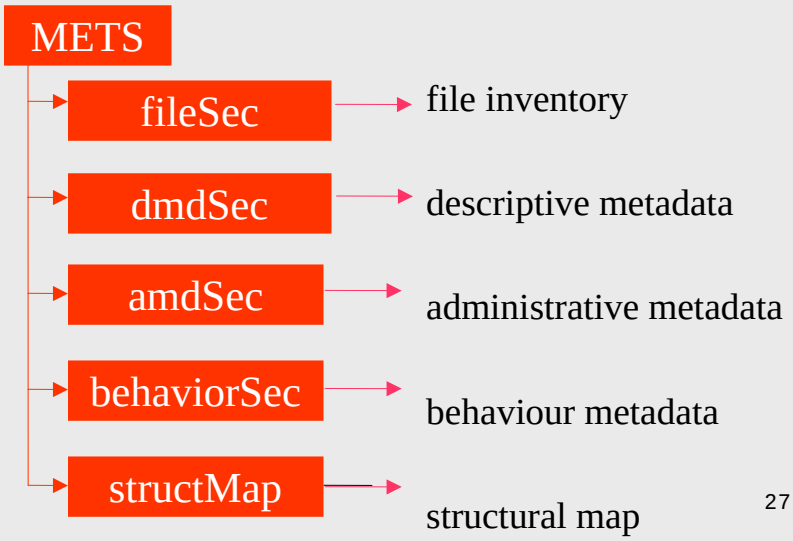
METS Sections

Defined in METS schema for navigation & browsing

1. Header (XML Namespaces)
2. File inventory,
3. Structural Map & Links
4. Descriptive Metadata (not part of METS but uses an externally developed descriptive metadata standard, e.g. MODS)
5. Administrative Metadata (points to external schemas):
    1. Technical, Source
    2. Digital Provenance
    3. Rights

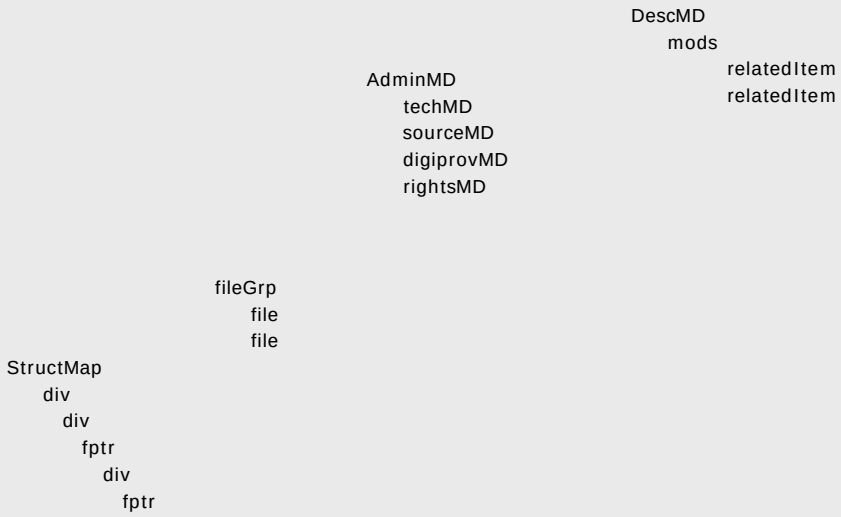
26

The structure of a METS file



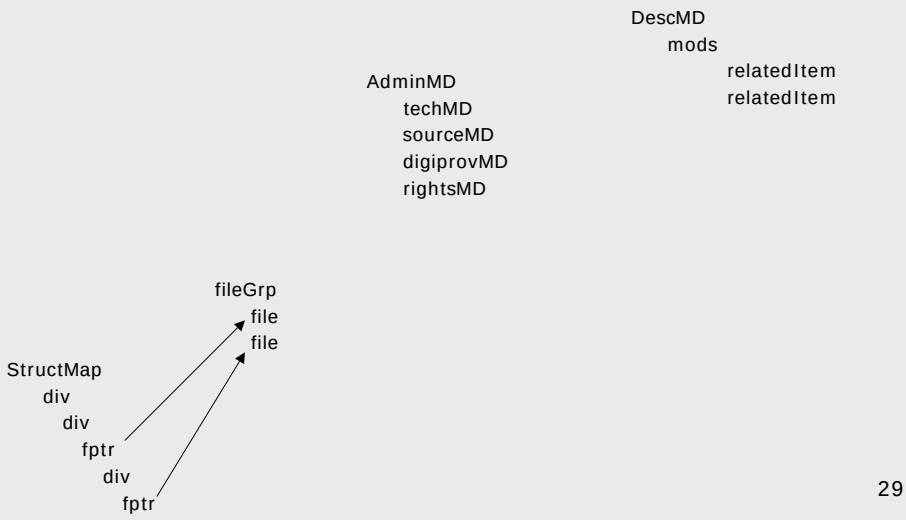
27

Linking in METS Documents
(XML ID / IDREF links)



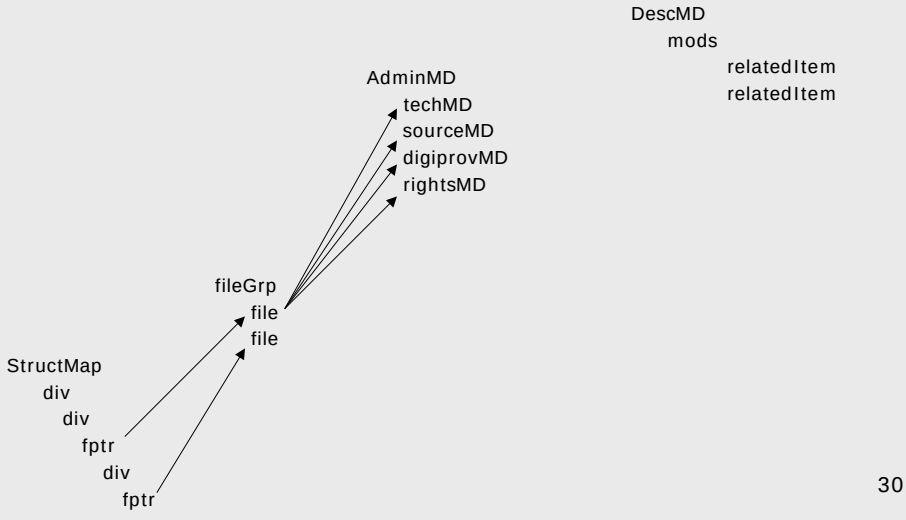
28

Linking in METS Documents
(XML ID/IDREF links)

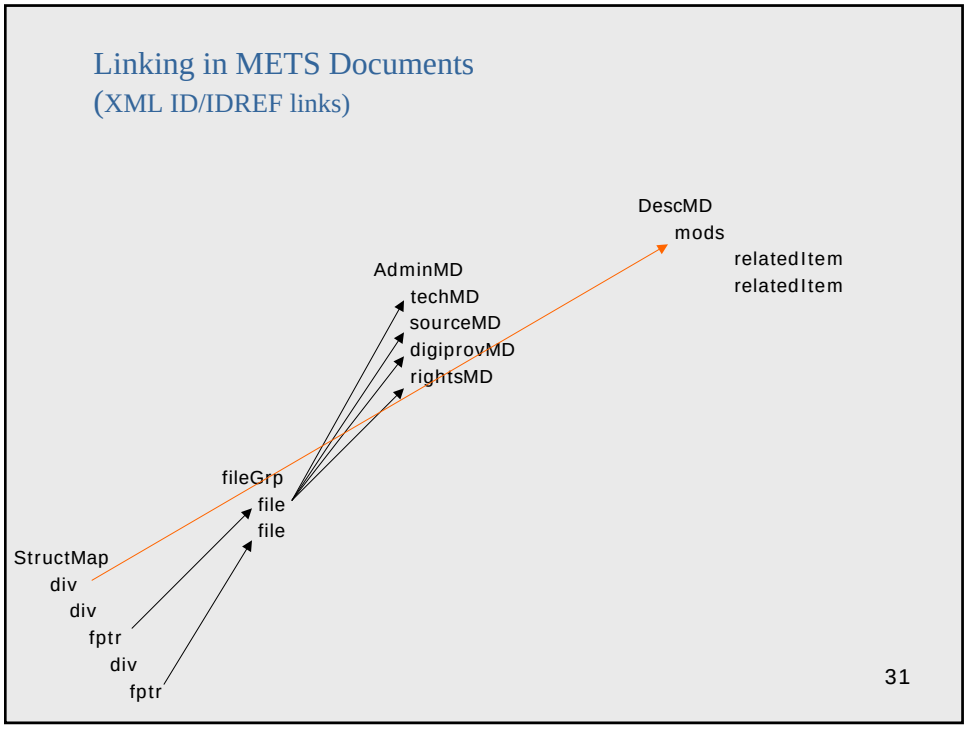


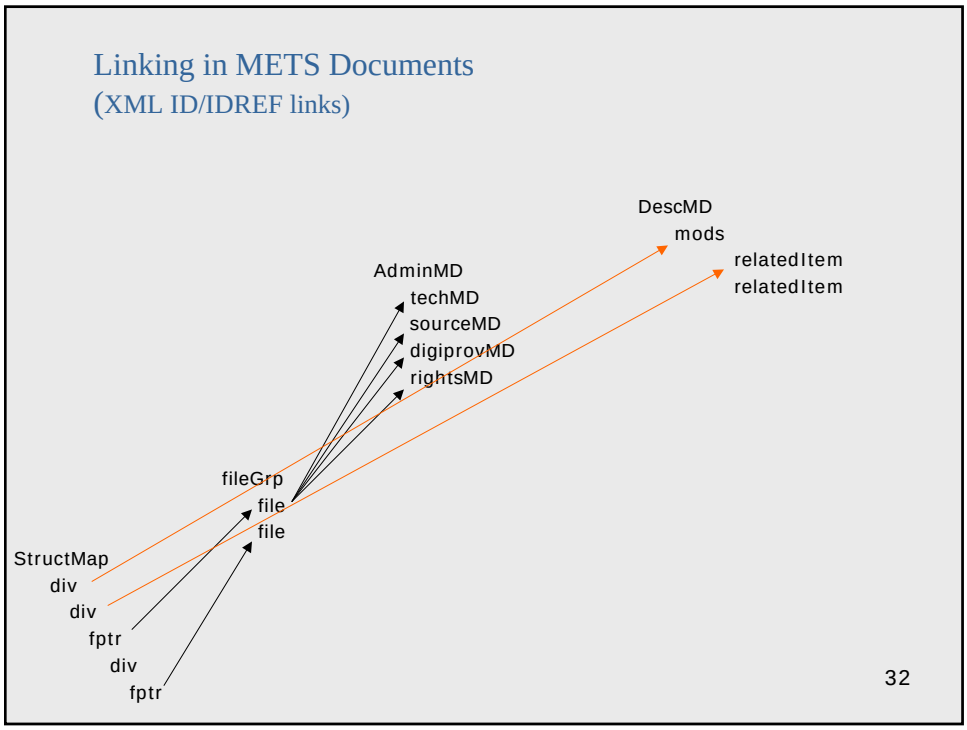
29

Linking in METS Documents
(XML ID/IDREF links)



30


Linking in METS Documents
(XML ID/IDREF links)
DescMD
mods
relatedItem
relatedItem
AdminMD
techMD
sourceMD
digiprovMD
rightsMD
fileGrp
file
file
StructMap
div
div
fptr
div
fptr
31


Linking in METS Documents
(XML ID/IDREF links)
DescMD
mods
relatedItem
relatedItem
AdminMD
techMD
sourceMD
digiprovMD
rightsMD
fileGrp
file
file
StructMap
div
div
fptr
div
fptr
32

METS extension schemas

- "wrappers" or "sockets" where elements from other schemas can be plugged in
- Provides extensibility
- Uses the XML Schema facility for combining vocabularies from different Namespaces
- Endorsed extension schemas:
  - Descriptive: MODS, DC, MARCXML
  - Technical metadata: MIX (image); textMD (text)
  - Preservation related: PREMIS

33

Descriptive Metadata Section (dmdSec)

Two methods: Reference and Wrap

```
<mets>
  <dmdSec></dmdSec>
  <fileSec></fileSec>
  <structMap></structMap>
</mets>
```

34

METS examples

- METS with MODS
  - Patriotic melodies
  - Recorded event
- METS with MODS, PREMIS and MIX
  - Portrait of Louis Armstrong (XML)
  - Portrait of Louis Armstrong (presentation)

35

MPEG-21 Digital Item Declaration (DID)

- ISO/IEC 21000-2: Digital Item Declaration
- A promising alternative to represent Digital Objects
- Starting to get supported by some repositories, e.g., aDORe, DSpace, Fedora
- A flexible and expressive model that easily represents compound objects (recursive “item”)
- Attach well-formed XML from persistent namespaces as metadata

36

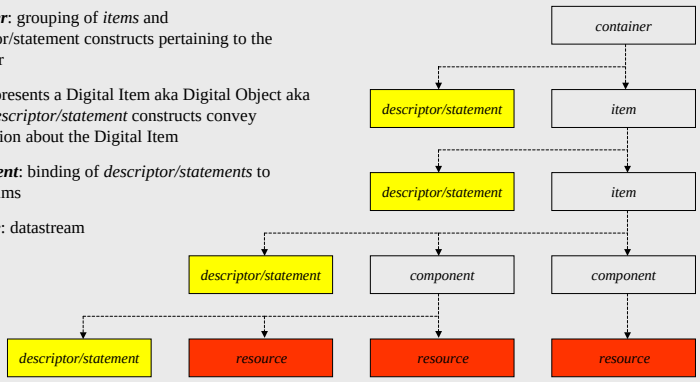
Abstract Model for MPEG-21 DID

container: grouping of *items* and descriptor/statement constructs pertaining to the container

item: represents a Digital Item aka Digital Object aka asset. *Descriptor/statement* constructs convey information about the Digital Item

component: binding of *descriptor/statements* to datastreams

resource: datastream



37

Summary: METS vs. MPEG 21 DID

- METS and MPEG DID are similar types of container formats in that both are expressed in XML, both represent the structure of digital objects, and both include metadata
- MPEG DID doesn't have the segmentation in metadata sections that METS does, so this implementation decision need not be made in DID
- METS is open source and developed by open discussion, mainly cultural heritage community
- MPEG DID is an ISO standard and has industry support, but is often implemented in a proprietary way and standards development is closed
- It would be possible to transform a METS container to a MPEG DID and vice versa; development of stylesheets will enable transformations

38

Exercise

- Encode your resource in DC and MODS using XML
- Use the template forms provided

39

Session 4: XML exercises

Use the following template for encoding the descriptive metadata in Dublin Core. Use simple Dublin Core elements but add attributes as needed.

```
<?xml version="1.0" encoding="UTF-8"?>
<metadata
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   http://dublincore.org/schemas/xmls/qdc/2003/04/02/dcterms.xsd
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:dcterms="http://purl.org/dc/terms/"
   xmlns:dcmitype="http://purl.org/dc/dcmitype/">

<!--  Begin Sample encoding  -->

   <dc:title xml:lang=" "> </dc:title>
   <dc:subject xsitype=" "> </dc:subject>

<!--  End Sample encoding  -->
```

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

```
</metadata>
```

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Session 4: XML exercises

Use the following template for encoding the descriptive metadata in MODS.

```
<?xml version="1.0" encoding="UTF-8"?>
<mods:mods>
<mods:titleInfo>
   <mods:title> </mods:title>
</mods:titleInfo>
<mods:name>
 <mods:namePart> </mods:namePart>
 <mods:role>
  <mods:roleTerm authority="marcrelator" type="text"> </mods:roleTerm>
 </mods:role>
</mods:name>
<mods:typeOfResource> </mods:typeOfResource>
<mods:originInfo>
   <mods:place>
    <mods:placeTerm> </mods:placeTerm>
  </mods:place>
  <mods:publisher> </mods:publisher>
  <mods:issuance> </mods:issuance>
</mods:originInfo>
<mods:physicalDescription>
 <mods:form authority="ihas">sheet music</mods:form>
 <mods:extent></mods:extent>
</mods:physicalDescription>
<mods:subject authority=" ">
<mods:topic></mods:topic>
<mods:topic> </mods:topic>
</mods:subject>
<mods:relatedItem type="host">
    <mods:titleInfo>
     <mods:title> </mods:title>
   </mods:titleInfo>
</mods:relatedItem>
<mods:location>
  <mods:physicalLocation authority="marcorg"> </mods:physicalLocation>
</mods:location>
<mods:recordInfo>
  <mods:recordContentSource>IHAS</mods:recordContentSource>
  <mods:recordChangeDate encoding="marc">060412</mods:recordChangeDate>
</mods:recordInfo>
</mods:mods>
```

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

5. Applying metadata standards: Application profiles

Metadata Standards and Applications Workshop

1

Goals of Session

- Learn how metadata standards are applied and used:
  - Learn about the concept and use of application profiles
  - Learn about how different metadata standards are used together in digital library applications
  - Understand questions for determining the appropriate metadata standard to use

2

Application profiles

- The set of metadata elements, policies, and guidelines defined for a particular application, implementation, or object type
- A declaration of the metadata terms an organization, information resource, application, or user community uses in its metadata
- Documents metadata standards used in instances, including schemas and controlled vocabularies, policies, required elements, etc.
- Called “application profile” or just “profile”

3

Why Application Profiles?

- Documenting user consensus
  - Usage
  - Obligation
- Many metadata standards are sufficiently flexible that they need a mechanism to impose some constraints
- “Mixing and matching metadata schemas”; documents how to use different metadata standards together
- Identifying appropriate vocabularies

4

Application profile components

- Human readable documentation
  - Define expectations for instances
  - Specific instructions for specific applications
  - Obligation and constraints
- Machine readable versions
  - May be processed by a machine, e.g. for validation of conformance
  - Based on RDF or XML
  - Should be based on a standard syntax

5

Dublin Core application profiles: term usages

- *Identifying attributes*
  - **Term URI, Name, Label, Defined By**
- *Definitional attributes*
  - **Definition, Comments, Type of Term**
- *Relational attributes*
  - **Refines, Refined By, Encoding Scheme For,**
  - **Uses Encoding Scheme, Similar To**
- *Constraints*
  - **Obligation, Condition, Datatype, Occurrence**

6

DCAP: mixing and matching metadata

- Currently a great deal of contention and discussion about the technical issues around re-use of properties
- Determining reusability of terms in a DCAP
  - Is the term a real "property" and defined as such within the source schema?
  - Is the term declared properly, with a URI and adequate documentation and support?
  - In general, properties whose meaning is partly or wholly determined by its place in a hierarchy are not appropriate for reuse without reference to the hierarchy.

7

Application profile development for RDA

- A DCMI/RDA Work Group is being formed to define RDA entities as an RDF vocabulary
- This will constitute an application profile of RDA according to the DCMI application profile framework
- RDA elements will be given URIs
- Controlled vocabularies in RDA will be represented in SKOS/RDF

8

What is a METS Profile?

- Description of a class of METS documents
  - provides document authors and programmers guidance to create and process conformant METS documents
- XML document using a schema
  - Expresses the requirements that a METS document must satisfy
- "Data standard" in its own right
  - A sufficiently explicit METS Profile may be considered a "data standard"
- METS Profiles are output in human-readable prose and not intended to be "machine actionable" (but they use a standard XML schema)

9

Components of a METS Profile

1. Unique URI
2. Short Title
3. Abstract
4. Date and time of creation
5. Contact Information
6. Related profiles
7. Extension schemas
8. Rules of description
9. Controlled vocabularies
10. Structural requirements
11. Technical requirements
12. Tools and applications
13. Sample document

10

MODS profiles

- Some applications are establishing MODS profiles to document usage, required elements, controlled vocabularies used, etc.

Some examples:

- DLF Aquifer MODS profile: to establish implementation guidelines for rich shared metadata for cultural heritage materials
- British Library electronic journal MODS profile

11

Using metadata standards together

- METS can be used to package together the metadata with the objects
  - MODS works well with METS for descriptive metadata and can be associated with any level of the description
  - Technical metadata can be inserted and associated with specific files

12

<dmdSec> with MODS Extension Schema

```
<mets:mets>
  …
  <mets:dmdSec>
    <mets:mdWrap>
      <mets:xmlData>
        <mods:mods></mods:mods>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:dmdSec>
  …
</mets:mets>
```

Descriptive metadata section

MODS data contained inside the metadata wrap section

Use of prefixes before element names to identify schema

13

<dmdSec> with <mods:relatedItem>

```
<mets:mets>
  …
  <mets:dmdSec>
    <mets:mdWrap>
      <mets:xmlData>
        <mods:mods>
          <mods:relatedItem type="constituent">
            <mods:relatedItem type="constituent"></mods:relatedItem>
          </mods:relatedItem>
        </mods:mods>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:dmdSec>
  …
</mets:mets>
```

The MODS relatedItem element can be nested and can be used to express a hierarchy.

14

```
<mods:mods>
   <mods:titleInfo>
      <mods:title>Bernstein conducts Beethoven </mods:title>
   </mods:titleInfo>
   <mods:name>
      <mods:namePart>Bernstein, Leonard</mods:namePart>
   </mods:name>
   <mods:relatedItem type="constituent">
      <mods:titleInfo>
         <mods:title>Symphony No. 5</mods:title>
      </mods:titleInfo>
      <mods:name>
         <mods:namePart>Beethoven, Ludwig van</mods:namePart>
      </mods:name>
         <mods:relatedItem  type="constituent">
            <mods:titleInfo>
               <mods:partName>Allegro con moto</mods:partName>
            </mods:titleInfo>
         </mods:relatedItem>
         <mods:relatedItem type="constituent">
            <mods:titleInfo>
               <mods:partName>Adagio</mods:partName>
            </mods:titleInfo>
         </mods:relatedItem>
   </mods:relatedItem>
</mods:mods>
```

15

MODS relatedItem *type="constituent"*

- Child element to MODS
- relatedItem element uses MODS content model
  - titleInfo, name, subject, physicalDescription, note, etc.
- Makes it possible to create rich analytics for contained works within a MODS record
- Repeatable and nestable recursively
  - Making it possible to build a hierarchical tree structure
- Makes it possible to associate descriptive data with any structural element

16

METS 2 Hierarchies: Logical & Physical

```
<mets:mets>
 <mets:dmdSec>
  <mets:mdWrap>
   <mets:xmlData>
    <mods:mods>
     <mods:relatedItem>
      <mods:relatedItem></mods:relatedItem>
     </mods:relatedItem>
    </mods:mods>
   </mets:xmlData>
  </mets:mdWrap>
 </mets:dmdSec>
 <mets:fileSec></mets:fileSec>
 <mets:structMap>
  <mets:div>
   <mets:div></mets:div>
  </mets:div>
 </mets:structMap>
</mets:mets>
```

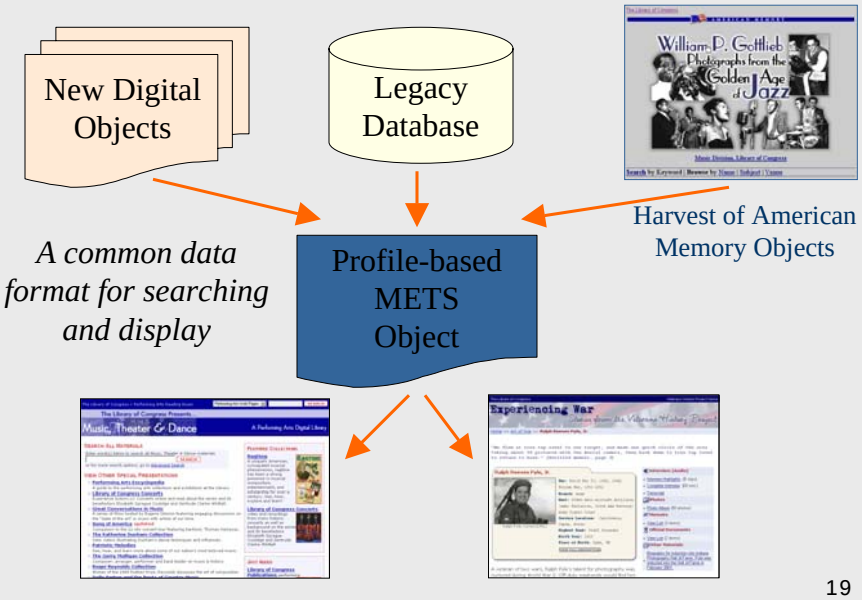
Hierarchy to represent "logical" structure (nested relatedItems)

Hierarchy to represent "physical" structure (nested div elements)

METS Profiles registered and used in applications (LC)

- Sheet Music
- Musical Score **(score, score and parts, or a set of parts only)**
- Print Material **(books, pamphlets, etc)**
- Music Manuscript **(score or sketches)**
- Recorded Event **(audio or video)**
- PDF Document
- Bibliographic Record
- Photograph
- Compact Disc
- Collection

Multiple Inputs to Common Data Format



19

Example: American Memory Harvest

- METS Photograph Profile
- William P. Gottlieb Collection
  Portrait of Louis Armstrong
- Photographic object

Convert file of 1600 MARC records, using *marc4j*, to XML modsCollection (single file).

Used XSLT stylesheet to create 1600 records conforming to the METS photograph profile.

Logical & Physical Relationships

Logical (MODS)

```
<mods:mods ID="ver01">
  <mods:titleInfo>
    <mods:title>Original Work</mods:title>
  </mods:titleInfo>
<mods:relatedItem type="otherVersion" ID="ver02">
    <mods:titleInfo>
      <mods:title>Derivative Work 2</mods:title>
    </mods:titleInfo>
</mods:relatedItem>
<mods:relatedItem type="otherVersion" ID="ver03">
    <mods:titleInfo>
      <mods:title>Derivative Work 3</mods:title>
    </mods:titleInfo>
</mods:relatedItem>
</mods:mods>
```

div TYPE="photo:version" *elements correspond to the 3 nodes using a logical sequence of ID to DMDID relationships*

Physical (METS structMap)

```
<mets:structMap>
  <mets:div TYPE="photo:photoObject"
            DMDID="MODS1">
    <mets:div TYPE="photo:version" DMDID="ver01">
      <mets:div TYPE="photo:image">
        <mets:fptr FILEID="FN10081"/>
      </mets:div>
    </mets:div>
    <mets:div TYPE="photo:version" DMDID="ver02">
      <mets:div TYPE="photo:image">
        <mets:fptr FILEID="FN10090"/>
      </mets:div>
      <mets:div TYPE="photo:version" DMDID="ver03">
        <mets:div TYPE="photo:image">
          <mets:fptr FILEID="FN1009F"/>
        </mets:div>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>
```

mods:mods *and* mods:relatedItem type ="otherVersion" *elements create a sequence of 3 nodes*

21

Advantages of METS/MODS Approach

- Ability to model complex library objects
- Ease of change and extension
  - both the data and the application
- Use of modern, non-proprietary software tools
- Use of XSLT for…
  - Legacy data conversion
  - Batch METS creation and editing
  - Web displays and behaviors
- Use of a common syntax – XML
  - For data creation, editing, storage and searching

continued…

22

Advantages of METS/MODS Approach

- Creation of multiple outputs from XML
  - HTML/XHTML for Web display; PDF for printing
- Ease of editing
  - Single records or selected batches of records
- Ability to validate data
- Ability to aggregate disparate data sources
- Ease of data management and publishing
- Excellent positioning for the future
  - New web applications (Web 2.0)
  - Repository submission and OAI harvesting
  - Cooperative projects (test interoperability)

23

Applying metadata standards: questions to ask for selection

- What type of material will be digitized?
- How rich does the metadata need to be?
- Is there information already available?
- Is there a Community of practice developed for this resource type(s)?
- What is the purpose of digital project?
- 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? Is there a need to interact with any existing records?
- What tools or systems options are available?

24

6. Applying metadata standards: Controlled vocabularies and quality issues

Metadata Standards and Applications Workshop

1

Goals of Session

- Understand how different controlled vocabularies are used in metadata
- Investigate metadata quality issues

2

Applying metadata: Controlled vocabularies

- Values that occur in metadata
- Often documented and published
- Goal to reduce ambiguity
- Control of synonyms
- Establishment of formal relationships among terms (where appropriate)
- Testing and validation of terms
- Metadata registries may be established

3

Why bother?

- To improve retrieval, i.e., to get an optimum balance of precision and recall
  - Precision – How many of the retrieved records are relevant?
  - Recall – How many of the relevant records did you retrieve?

4

Types of Controlled Vocabularies

- Lists of enumerated values
- Taxonomy
- Thesaurus
- Classification Schemes
- Ontology

5

Lists

A list is a simple group of terms

Example:

Alabama
Alaska
Arkansas
California
Colorado
. . . .

Frequently used in Web site pick lists and pull down menus

6

Taxonomies

A *taxonomy* is a set of preferred terms, all connected by a hierarchy or polyhierarchy

Example:

Chemistry
Organic chemistry
Polymer chemistry
Nylon

Frequently used in web navigation systems

7

Thesauri

A *thesaurus* is a controlled vocabulary with multiple types of relationships

Example:

Rice
UF paddy
BT Cereals
BT Plant products
NT Brown rice
RT Rice straw

8

Ontology

- One definition: "An arrangement of concepts and relations based on an underlying model of reality."
  - Ex.: Organs, symptoms, and diseases in medicine
- No real agreement on definition—every community uses the term in a slightly different way

9

Thesaurus Relationships

Relationship types:

- Use/Used For – indicates preferred term
- Hierarchy – indicates broader and narrower terms
- Associative – almost unlimited types of relationships may be used

It is the most complex format for controlled vocabularies and widely used.

10

ANSI/NISO Z39.19-2005 ISBN: 1-880124-65-3

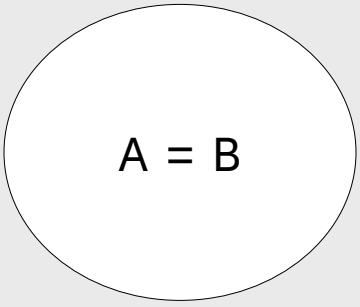
Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

Abstract: This Standard presents guidelines and conventions for the contents, display, construction, testing, maintenance, and management of monolingual controlled vocabularies. This Standard focuses on controlled vocabularies that are used for the representation of content objects in knowledge organization systems including lists, synonym rings, taxonomies, and thesauri. This Standard should be regarded as a set of recommendations based on preferred techniques and procedures. Optional procedures are, however, sometimes described, e.g., for the display of terms in a controlled vocabulary. The primary purpose of vocabulary control is to achieve consistency in the description of content objects and to facilitate retrieval. Vocabulary control is accomplished by three principal methods: defining the scope, or meaning, of terms; using the equivalence relationship to link synonymous and nearly synonymous terms; and distinguishing among homographs.

11

Equivalence Relationships

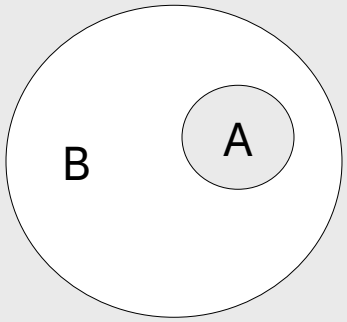
Term A and Term B overlap completely



12

Hierarchical Relationships

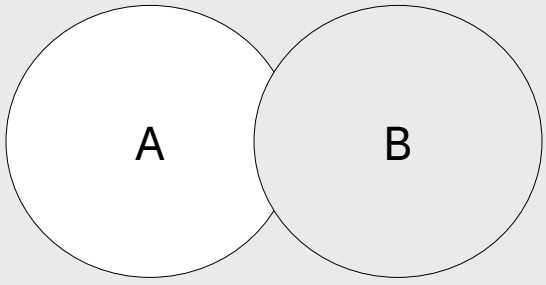
- Term A is included in Term B



13

Associative Relationships

- Semantics of terms A and B overlap



14

Expressing Relationship

Relationship	Rel. Indicator	Abbreviation
Equivalence (synonymy)	Use Used for	None or U UF
Hierarchy	Broader term Narrower term	BT NT
Association	Related term	RT

15

Vocabulary Management

- The degree of control over a vocabulary is (mostly) independent of its type
  - **Uncontrolled** – Anybody can add anything at any time and no effort is made to keep things consistent
  - **Managed** – Software makes sure there is a list that is consistent (no duplicates, no orphan nodes) at any one time. Almost anybody can add anything, subject to consistency rules
  - **Controlled** – A documented process is followed for the update of the vocabulary. Few people have authority to change the list. Software may help, but emphasis is on human processes and custodianship

16

Encoding controlled vocabularies

- MARC 21
  - Authority Format used for names, subjects, series
  - Classification Format used for formal classification schemes
- MADS (a derivative of MARC)
- Simple Knowledge Organization System (SKOS)
  - Intended primarily for concept schemes (e.g., not names)

17

SKOS

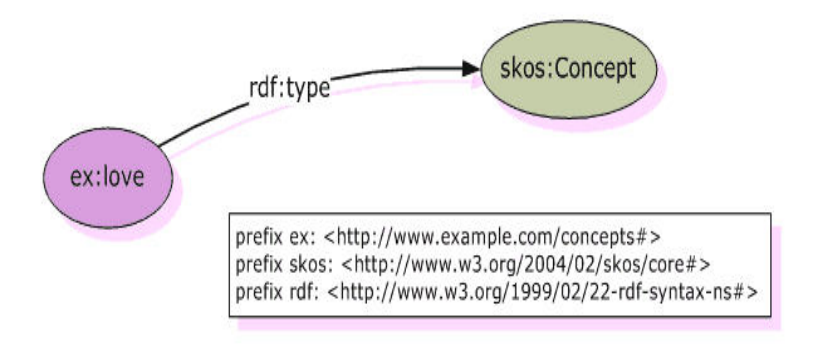
- "SKOS Core provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies', other types of controlled vocabulary, and also concept schemes embedded in glossaries and terminologies."

--SKOS Core Guide

18

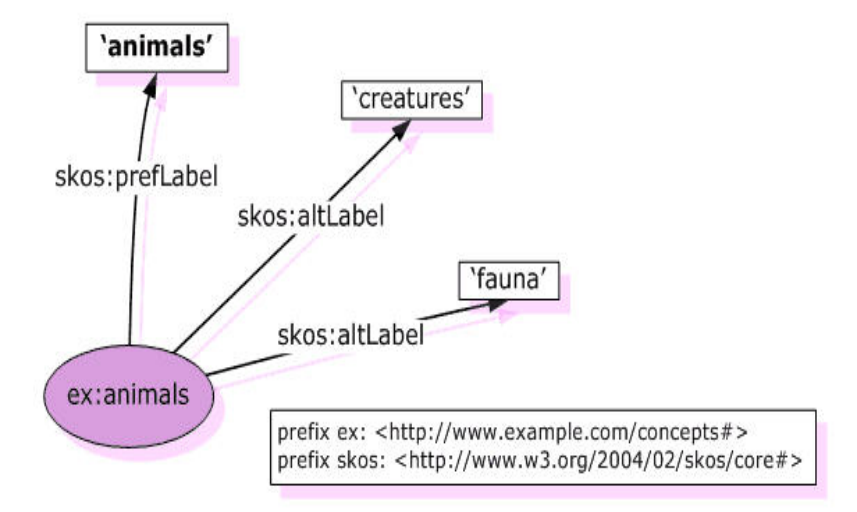
The skos:Concept class allows you to assert that a resource is a *conceptual resource*.

That is, the resource *is itself* a concept.



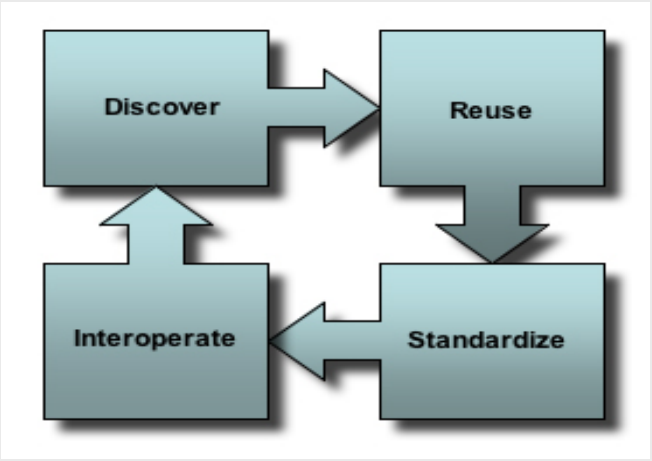
19

Preferred and Alternative Lexical Labels



20

Registries: the Big Picture



(Adapted from Wagner & Weibel, "The Dublin Core Metadata Registry: Requirements, Implementation, and Experience" JoDI, 2005)

21

Why Registries?

- Support interoperability
  - Discovery of available schemes and schemas for description of resources
  - Promote reuse of extant schemes and schemas
  - Access to machine-readable and human-readable services
  - Support for crosswalking and translation
- Coping with difference metadata schemes

22

Declaration, documentation, publication

- To identify the source of a vocabulary, e.g., a term comes from LCSH, as identified in my metadata by a URI: info:lcsh
- To clarify a term and its definition
- To publish controlled vocabularies and have access to information about each term

23

Some uses for registries

- Metadata Schemas
  - Crosswalks between metadata schemas
- Controlled Vocabularies
  - Mappings between vocabularies
- Application Profiles
  - Schema and vocabulary information in combination

24

Metadata registries

- Some are formal, others are informal lists
- Some formal registries:
  - Dublin Core registry of DC terms
  - NSDL registry of vocabularies used
  - LC is establishing registries
    - MARC and ISO code lists
    - Enumerated value lists
    - LCSH in SKOS

25

Applying metadata standards: quality issues

- Defining quality
- Criteria for assessing quality
- Levels of quality
- Quality indicators

26

Determining and Ensuring Quality

- What constitutes quality?
- Techniques for evaluating and enforcing consistency and predictability
- Automated metadata creation: advantages and disadvantages
- Metadata maintenance strategies

27

Quality Measurement: Criteria

- Completeness
- Accuracy
- Provenance
- Conformance to expectations
- Logical consistency and coherence
- Timeliness (Currency and Lag)
- Accessibility

28

Basic Quality Levels

- Semantic structure ("format," "schema" or "element set")
- Syntactic structure (administrative wrapper and technical encoding)
- Data values or content

29

Quality Indicators: Tier 1

- Technically valid
  - Defined technical schema; automatic validation
- Appropriate namespace declarations
  - Each element defined within a namespace; not necessarily machine-resolvable
- Administrative wrapper present
  - Basic provenance (unique identifier, source, date)

30

Quality Indicators: Tier 2

- Controlled vocabularies
  - Linked to publicly available sources of terms by unique tokens
- Elements defined and documented by a specific community
  - Preferably an available application profile
- Full complement of general elements relevant to discovery
- Provenance at a more detailed level
  - Methodology used in creation of metadata?

31

Quality Indicators: Tier 3

- Expression of metadata intentions based on documented AP endorsed by a specialized community and registered in conformance to a general metadata standard
- Source of data with known history of updating, including updated controlled vocabularies
- Full provenance information (including full source info), referencing practical documentation

32

Improving Metadata Quality ...

- Documentation
  - Basic standards, best practice guidelines, examples
  - Exposure and maintenance of local and community vocabularies
  - Application Profiles
  - Training materials, tools, methodologies

33

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

7. Approaches to Models of Metadata Creation, Storage and Retrieval

Metadata Standards
and Applications

1

Goals of Session

- Understand the differences between traditional vs. digital library Metadata Creation
- Storage, and
- Retrieval/Discovery

2

Creating Metadata Records

- The "Library Model"
  - Trained catalogers, one-at-a-time metadata records
- The "Submission Model"
  - Authors create metadata when submitting resources
- The "Automated Model"
  - Automated tools create metadata for resources
- "Combination Approaches"

3

The Library Model

- Records created "by hand," one at a time
- Shared documentation and content standards (AACR2, etc.)
- Efficiencies achieved by sharing information on commonly held resources
- Not easily extended past the granularity assumptions in current practice

4

The Submission Model

- Based on author or user generated metadata
- Can be wildly inconsistent
  - Submitters generally untrained
  - May be expert in one area, clueless in others
- Often requires editing support for usability
- Inexpensive, may not be satisfactory as an only option

5

The Automated Model

- Based largely on text analysis; doesn't usually extend well to non-text or low-text
- Requires development of appropriate evaluation and editing processes
- Still largely research; few large, successful production examples, yet
- Can be done in batch
- Also works for technical as well as descriptive metadata

6

Combination Approaches

- Combination machine and human created metadata
  - Ex.: LC Web Archives (http://loc.gov/minerva)
- Combination metadata and content indexing
  - Ex.: NSDL (http://nsdl.org)

7

Content "Storage" Models

- 'Storage' in this context relates to the relationships between metadata and content (not the systems that purport to 'store' content for various uses)
- These relationships affect how access to the information is accomplished, and how the metadata either helps or hinders the process (or is irrelevant to it)

8

Common ‘Storage’ Models

- Content with metadata
- Metadata only
- Service only

9

Content with metadata

- Examples:
  - HTML pages with embedded ‘meta’ tags
  - Most content management systems (though they may store only technical or structural metadata)
  - Text Encoding Initiative (TEI)
- Often proves difficult to update

10

Metadata only

- Library catalogs
  - Web-based catalogs often provide some services for digital content
- Electronic Resource Management Systems (ERMS)
  - Provide metadata records for title level only
- Metadata aggregations
  - Using API or OAI-PMH for harvest and re-distribution

11

Service only

- Often supported partially or fully by metadata
  - Google, Yahoo (and others)
    - Sometimes provide both search services and distributed search software
  - Electronic journals (article level)
    - Linked using 'link resolvers' or available independently from websites
    - Have metadata behind their services but don't generally distribute it separately

12

Common Retrieval Models

- Library catalogs
- Web-based ("Amazoogle")
- Portals and federations

13

Library Catalogs

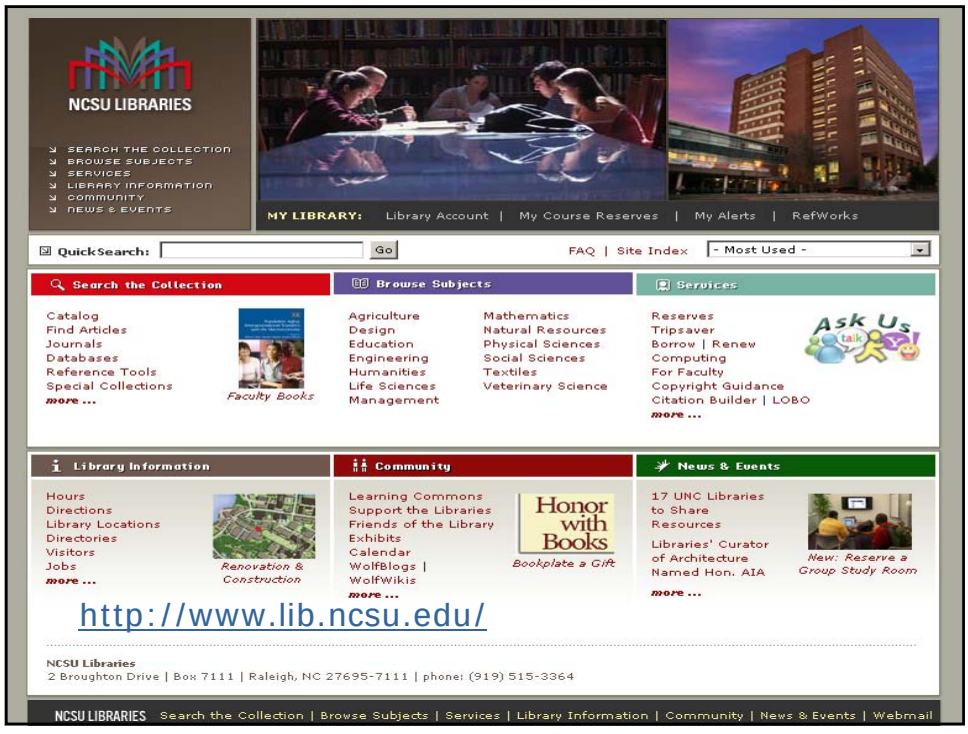
- Based on a consensus that granular metadata is useful
- Expectations of uniformity of information content and presentation
- Designed to optimize recall and precision
- Addition of relevance ranking and keyword searching of limited value (only 'text' used is the metadata itself)
- Retrieval options limited by LMS vendor decisions

14

Endeca

- Information Access Platform (software)
  optimizes huge volumes of data and can be used by everyday people
- North Carolina State University Libraries
  2006, first library to use this technology
  Fast, flexible & as simple as browsing the Web, but unmatched by Web search engines due to rich [metadata] content & cutting-edge capabilities.

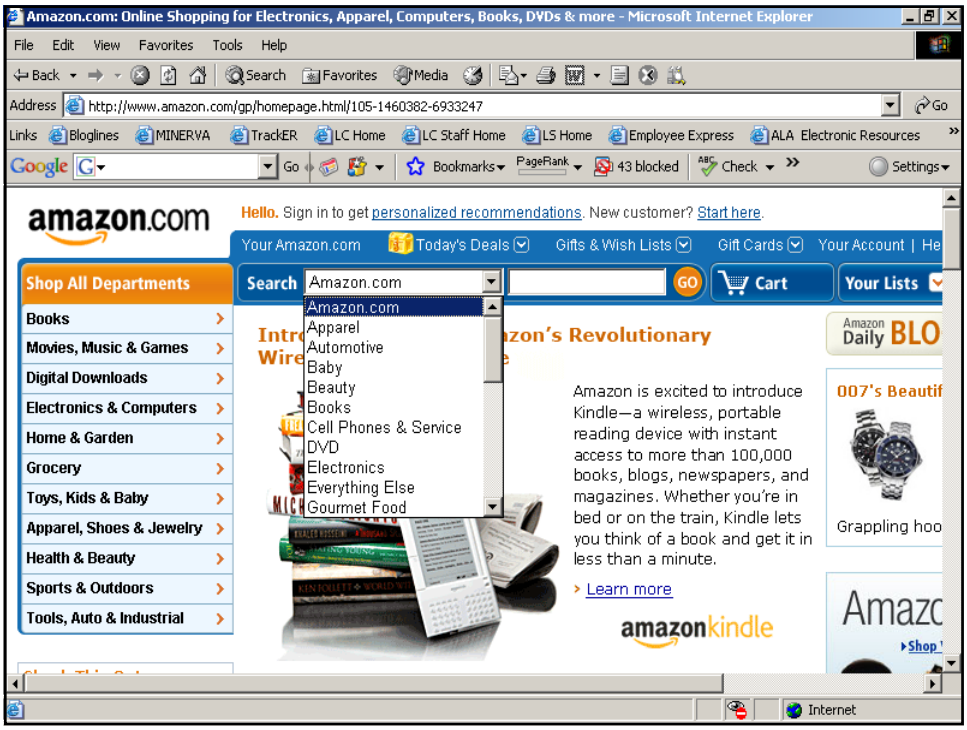
15



Web-based

- The “Amazoogle” model:
  - Lorcan Dempsey: “Amazon, Google, eBay: massive computational and data platforms which exercise strong gravitational web attraction.”
  - Based primarily on full-text searching and link- or usage-based relevance ranking (lots of recall, little precision)
  - Some efforts to combine catalog and Amazoogle searches (ex.: collaborations with WorldCat)
  - Google is using metadata

17



Portals and Federations

- **Portals**: defined content boundaries
  - Some content also available elsewhere
  - ex.: Specific library portals, subject portals like Portals to the World (http://www.loc.gov/rr/international/portals.html)
- **Federations**: protected content and services
  - Often specialized services based on specifically purposed metadata (ex.: BEN- http://www.biosciednet.org/portal/)

19

Information Discovery and Retrieval

- Z39.50
- SRU
- Federated search (Metasearch)

20

Federated search

- Institutions are experimenting with Federated search (meta-search) to search across disparate sources of metadata
- LC has a beta version available:

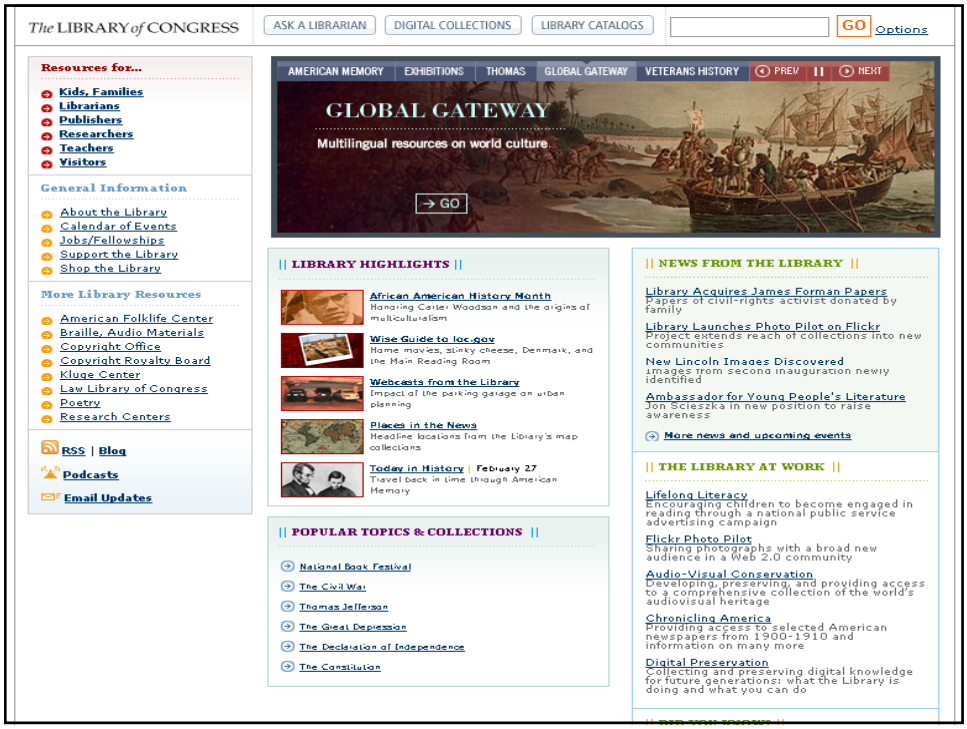
http://www.loc.gov/search/new/

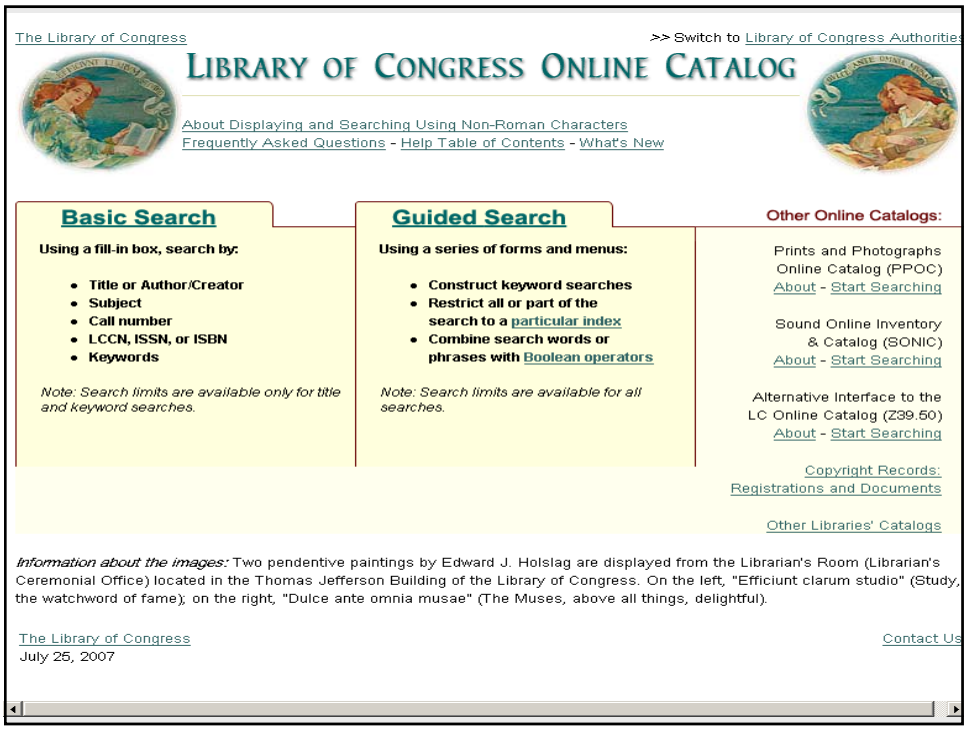
21

Z39.50

- An international (ISO 23950) standard defining a protocol for computer-to-computer information retrieval.
- Makes it possible for a user in one system to search and retrieve information from other computer systems (that have also implemented Z39.50)
- Originally approved by the National Information Standards Organization (NISO) in 1988

22


The LIBRARY of CONGRESS
ASK A LIBRARIAN
DIGITAL COLLECTIONS
LIBRARY CATALOGS
GO
Options
Resources for...
Kids, Families
Librarians
Publishers
Researchers
Teachers
Visitors
General Information
About the Library
Calendar of Events
Jobs/Fellowships
Support the Library
Shop the Library
More Library Resources
American Folklife Center
Braille, Audio Materials
Copyright Office
Copyright Royalty Board
Kluge Center
Law Library of Congress
Poetry
Research Centers
RSS | Blog
Podcasts
Email Updates
AMERICAN MEMORY
EXHIBITIONS
THOMAS
GLOBAL GATEWAY
VETERANS HISTORY
PREV
NEXT
GLOBAL GATEWAY
Multilingual resources on world culture
GO
|| LIBRARY HIGHLIGHTS ||
African American History Month
Hanging Carter Woodson and the origins of multiculturalism
Wise Guide to loc.gov
Home movies, stinky cheese, Denmark, and the Main Reading Room
Webcasts from the Library
Impact of the parking garage on urban planning
Places in the News
Headline locations from the Library's map collections
Today in History | February 27
Travel back in time through American Memory
|| POPULAR TOPICS & COLLECTIONS ||
National Book Festival
The Civil War
Thomas Jefferson
The Great Depression
The Declaration of Independence
The Constitution
|| NEWS FROM THE LIBRARY ||
Library Acquires James Forman Papers
Papers of civil-rights activist donated by family
Library Launches Photo Pilot on Flickr
Project extends reach of collections into new communities
New Lincoln Images Discovered
Images from second inauguration newly identified
Ambassador for Young People's Literature
Jon Scieszka in new position to raise awareness
More news and upcoming events
|| THE LIBRARY AT WORK ||
Lifelong Literacy
Encouraging children to become engaged in reading through a national public service advertising campaign
Flickr Photo Pilot
Sharing photographs with a broad new audience in a Web 2.0 community
Audio-Visual Conservation
Developing, preserving, and providing access to a comprehensive collection of the world's audiovisual heritage
Chronicling America
Providing access to selected American newspapers from 1900-1910 and information on many more
Digital Preservation
Collecting and preserving digital knowledge for future generations: what the Library is doing and what you can do


The Library of Congress
>> Switch to Library of Congress Authorities
LIBRARY OF CONGRESS ONLINE CATALOG
About Displaying and Searching Using Non-Roman Characters
Frequently Asked Questions - Help Table of Contents - What's New
Basic Search
Using a fill-in box, search by:
Title or Author/Creator
Subject
Call number
LCCN, ISSN, or ISBN
Keywords
Note: Search limits are available only for title and keyword searches.
Guided Search
Using a series of forms and menus:
Construct keyword searches
Restrict all or part of the search to a particular index
Combine search words or phrases with Boolean operators
Note: Search limits are available for all searches.
Other Online Catalogs:
Prints and Photographs Online Catalog (PPOC)
About - Start Searching
Sound Online Inventory & Catalog (SONIC)
About - Start Searching
Alternative Interface to the LC Online Catalog (Z39.50)
About - Start Searching
Copyright Records: Registrations and Documents
Other Libraries' Catalogs
Information about the images: Two pendentive paintings by Edward J. Holslag are displayed from the Librarian's Room (Librarian's Ceremonial Office) located in the Thomas Jefferson Building of the Library of Congress. On the left, "Efficiunt clarum studio" (Study, the watchword of fame); on the right, "Dulce ante omnia musae" (The Muses, above all things, delightful).
The Library of Congress
July 25, 2007
Contact Us

Simple Search

Z39.50 Gateway to the
Library of Congress Online Catalog

Enter Search Term(s):
(The search term can be a single word or a phrase from anywhere in the record. Enter an author's name in indirect order, i.e., last_name, first_name.)

Submit Query | Clear Form

Use of this form results in a search of the LC Voyager database (approximately 14 million records). This database contains records in all bibliographic formats (i.e., books, serials, music, maps, manuscripts, computer files, and visual materials), and includes the retrospective, unedited older bibliographic records known as the PreMARC File. LC name and subject authority records cannot be searched.

For more complex searches, use Advanced Search (multiple terms using Boolean operators)
For phrase searches, use Left-Anchored Phrase Search
To search a different database, use LC WWW/Z39.50 Gateway

Library of Congress
Contact Us (April 29, 2003)

25

International Standard Maintenance Agency **Z39.50** The Library of Congress Network Development & MARC Standards Office

Z39.50 Resources - **Z39.50 Document** - **Related Specifications** - **Object Identifiers**
Implementor Register - **Z39.50 Profiles** - **ZIG Meetings** - **Site Index** - **ZIG Listserv**

This page provides links to information about Z39.50 resources and about the development and maintenance of Z39.50 (existing as well as future versions) and the implementation and use of the Z39.50 protocol.

"Z39.50" refers to the International Standard, ISO 23950: "Information Retrieval (Z39.50): Application Service Definition and Protocol Specification", and to ANSI/NISO Z39.50. The Library of Congress is the Maintenance Agency and Registration Authority for both standards, which are technically identical (though with minor editorial differences).

The standard specifies a client/server-based protocol for searching and retrieving information from remote databases.

Z39.50 Proposals

Z39.50 Maintenance Agency: z3950@loc.gov
Maintenance Agency Procedures

See also:
SRU (Search/Retrieve via URL) -- a standard search protocol for Internet search queries.
CQL (Common Query Language) -- standard query syntax for representing queries.

They are based on Z39.50 semantics.

Library of Congress Home - Other Standards Maintained by the Library - Z39.50 Gateway

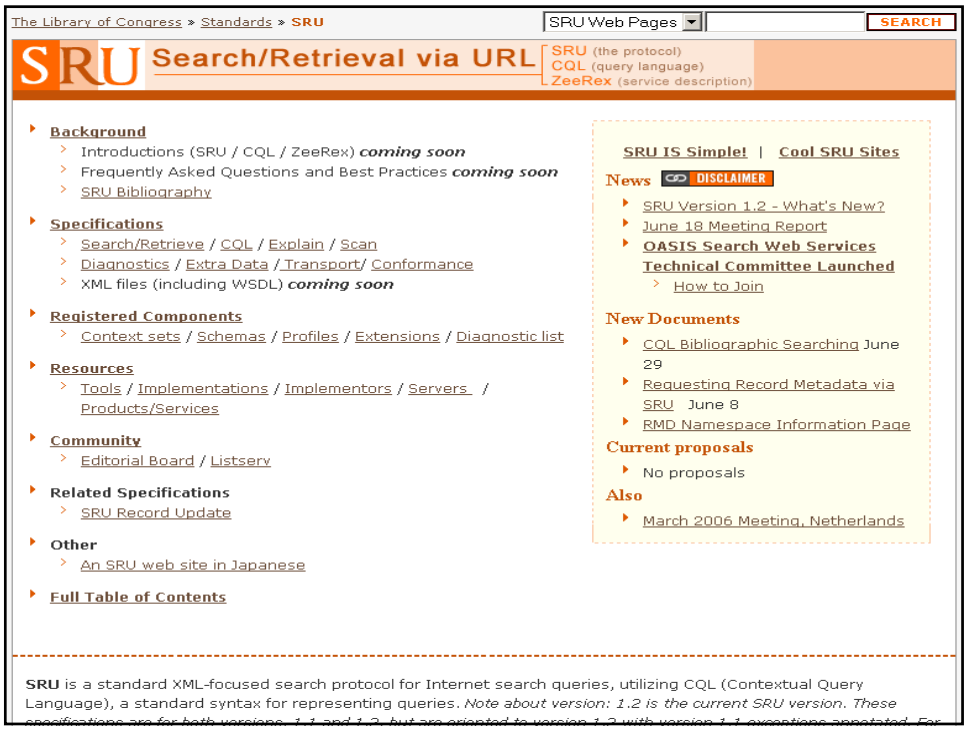
Contact us
Updated: October 1, 2007

26

SRU
Search/Retrieval via URL

- **SRU** is a standard XML-focused search protocol for Internet search queries, utilizing CQL (Contextual Query Language), a standard syntax for representing queries
- To learn about it see: http://www.loc.gov/standards/sru/index.html

27



Can You Tell?

- Can you tell what's going on behind these sites?
- How are they organized?
- What creation and storage models are used?
  - *Chopin Early Editions:* http://chopin.lib.uchicago.edu/
  - *The Great Kanto Earthquake of 1923:* http://dl.lib.brown.edu/kanto/
  - *Mark Twain Project:* http://marktwainproject.org

29

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

8. Metadata Interoperability and Crosswalks

Metadata Standards and Applications Workshop

1

Goals of Session

- Understand interoperability protocols (OAI-PMH, OpenURL for reference)
- Understand crosswalking and mapping as it relates to interoperability

2

Tools For Sharing Metadata/Interoperability

- Protocols
  - OAI-PMH for harvesting
  - OpenURL for reference linking
- Good practices and documentation
- Crosswalking

3

What's the Point of Interoperability?

- For users, it's about resource discovery (user tasks)
  - What's out there?
  - Is it what I need for my task?
  - Can I use it?
- For resource creators, it's about distribution and marketing
  - How can I increase the number of people who find my resources easily?
  - How can I justify the funding required to make these resources available?

4

OAI-PMH

- Open Archives Initiative Protocol for Metadata Harvesting (http://www.openarchives.org/)
- Roots in the ePrint community, although applicability is much broader
- Mission: "The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content."
- Content in this context is actually "metadata about content"

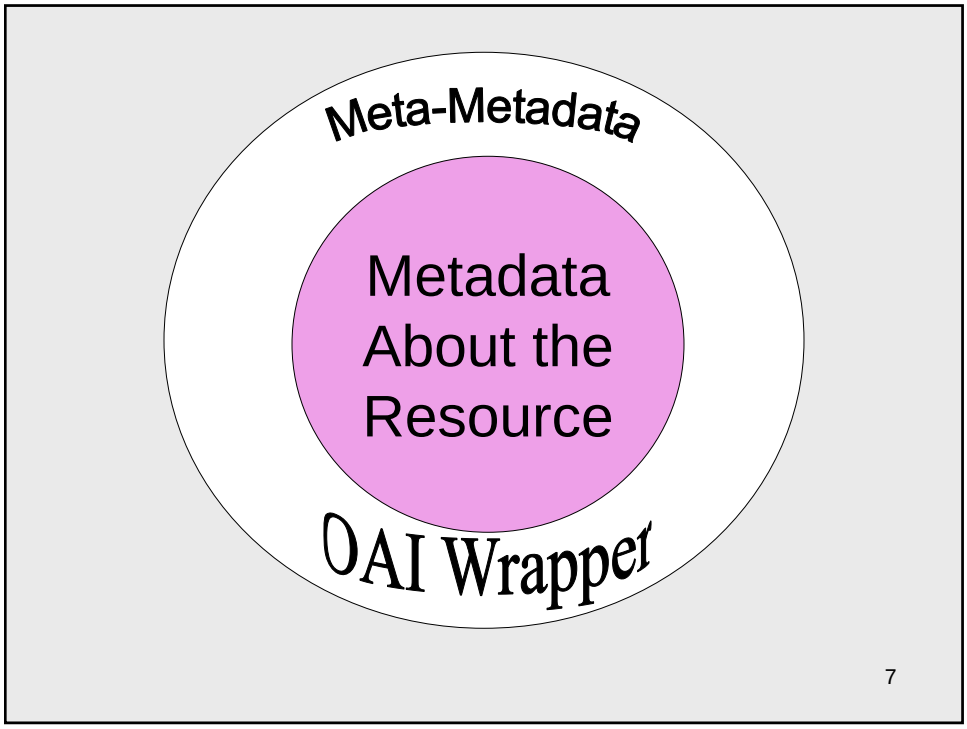
5

OAI-PMH in a Nutshell

- Essentially provides a simple protocol for "harvest" and "exposure" of metadata records
- Specifies a simple "wrapper" around metadata records, providing metadata about the record itself
- OAI-PMH has been about the metadata, not about the resources

  ARTstor cdwa-Lite experiment http://www.artstor.org/index.shtml

6



What was OAI-PMH designed for?

- Way to distribute records to other libraries
- Low barrier to entry for record providers
- Based on
  - Records must be in XML
  - OAI-PMH supports any metadata format encoded in XML—Simple Dublin Core is the minimal format specified
- Not Z39.50
  - Not a way to support federated search
  - No “on-the-fly” sets.
- More like CDS service, but it’s free,
  - users “pull” records when they want, at intervals that are convenient for them (every day, every hour, on any schedule, or ad hoc)

8

OAI-PMH: Data Provider

- Has records to share
- Runs system that responds to requests
  - following protocol
- Advertises base URL from which records are harvestable
- Just leaves system running
  - **No human intervention needed to service requests**
  - **Can control level of activity to protect performance for primary users**

9

OAI-PMH: Service Provider

- Assumed to be providing “union catalog” service
  - OAIster { http://www.oaister.org/ }
- or a specialist, value-added service
  - Sheet Music Consortium {http://digital.library.ucla.edu/sheetmusic/ }
- Harvests records, with ability to select limited to
  - Records updated in a certain timespan
  - Predetermined sets of records (like CDS)
  - Known records by identifiers (OAI identifiers, not LCCNs)

10

Inside OAI Repositories

- *repository* - A repository is a network accessible server that can process requests. A repository is managed by a data provider to expose metadata to harvesters
- *resource* - A resource is the object or "stuff" that metadata is "about," whether physical or digital, stored in the repository or a constituent of another database
- *item* - An item is a constituent of a repository from which metadata about a resource can be disseminated.
- *record* - A record is metadata in a specific metadata format

11

Protocol has 6 requests

- Identify
  - Facts about the data provider service
- ListMetadataFormats
- ListSets
  - Predetermined sets of records
- ListIdentifiers
  - Refine by set, date range for last update
  - Good way to count records
- GetRecord
  - By record identifier
- ListRecords
  - Like ListIdentifiers but with full records in specified format

More info at http://www.openarchives.org/

12

http://www.oaforum.org/tutorial/english/page3.htm#section3

13

OAI at LC

- LC as OAI Data Provider for historical collections
  - http://memory.loc.gov/ammem/oamh/lcoa1_content.html
  - Adding new collections steadily
  - MARC source (so far)
  - Handles rather than regular URLs in 856 $u
  - Records upgraded from PREMARC (minimally, not AACR2)
  - Records available as MARCXML, MODS or DC
- *Available and Useful: OAI at the Library of Congress*
  - Library Hi Tech, Volume 21, No. 2, 2003, pp. 129-139 [DOI:10.1108/07378830310491899]
  - http://memory.loc.gov/ammem/techdocs/libht2003.html

14

A Simple OAI Example

```
<?xml version="1.0" encoding="UTF-8"?>
<oai_dc:dc
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:title xml:lang="en">Grassmann's space analysis</dc:title>
  <dc:creator>Hyde, E. W. (Edward Wyllys)</dc:creator>
  <dc:subject>LCSH:Ausdehnungslehre; LCCN QA205.H99</dc:subject>
  <dc:publisher>J. Wiley &amp;amp; Sons</dc:publisher>
  <dc:date>Created: 1906; Available: 1991</dc:date>
  <dc:type>text</dc:type>
  <dc:identifier>http://resolver.library.cornell.edu/math/1796949
     </dc:identifier>
  <dc:language>english</dc:language>
  <dc:rights xml:lang="en">Public Domain</dc:rights>
</oai_dc:dc>
```

15

Demo:

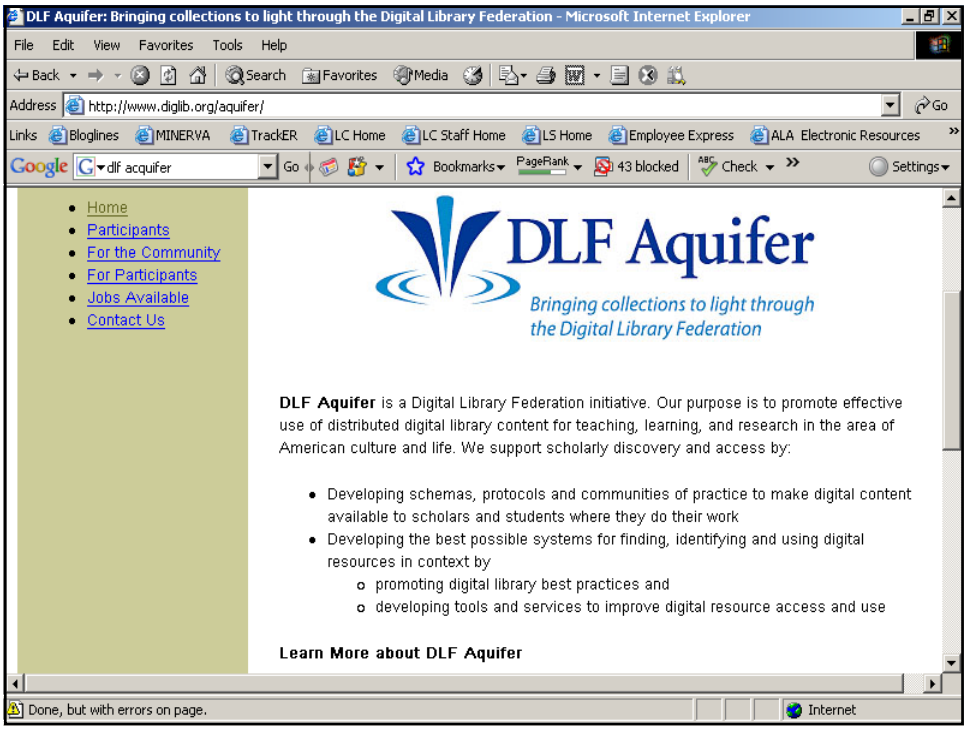
- Straight from browser
  - Sample queries on LC site
    - http://memory.loc.gov/ammem/oamh/
- Harvesters need to:
  - Work with XML
  - Pick a metadata format
  - Understand how to harvest complete set
- Could easily give a bit more help online
  - More or different canned options (but not to harvest whole set)
  - Explanation of how to harvest a whole set
  - Point to OAI-harvesting software libraries

16

OAI Best Practices Activities

- Sponsored by Digital Library Federation (DLF)
- Guidelines for data providers and service providers
  - http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl
  - Not just DLF, also NSDL
  - Best Practices for Shareable Metadata
    - http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl?PublicTOC
- Workshops to encourage DLF members to make records for their digitized content harvestable
  - Also sponsored by IMLS

17



OAIster

http://www.oaister.org/

- A **union catalog of digital resources**. Provides access to digital resources by "harvesting" their descriptive metadata (records) using OAI-PMH.
- Currently provides access to 14,900,092 records from 939 contributors.

19

What's an OpenURL?

- The OpenURL provides a standardized format for transporting bibliographic metadata about objects between information services
- Provides a basis for building services via the notion of an *extended service-link*, which moves beyond the classic notion of a *reference link* (a link from metadata to the full-content described by the metadata)

21

OpenURL Characteristics

- Protocol operates between an information resource and a service component
- Service component is called a "link server" or "link resolver"
- Link server defines the user context
- Takes source citation and determines whether a user has access

22

Distinguishing Users

- Uses information stored in a cookie (the CookiePusher mechanism)
- Uses information contained in a digital certificate, such as the one proposed by the DLF digital certificates prototype project
- Identifies a user's IP address
- Obtains user attributes via the Shibboleth framework

23

Additional Open URL Services

- Link from a record in an abstracting and indexing database (A&I) to the full-text described by the record
- Link from a record describing a book in a library catalogue to a description of the same book in an Internet book shop
- Link from a reference in a journal article to a record matching that reference in an A&I database
- Link from a citation in a journal article to a record in a library catalogue that shows the library holdings of the cited journal

24

OpenURL Examples & Demo

- http://sfxserver.uni.edu/sfxmenu?issn=1234-5678&date=1998&volume=12&issue=2&spage=134
- An OpenURL demo:
  - http://www.ukoln.ac.uk/distributed-systems/openurl/

25

Crosswalking

"Crosswalks support conversion projects and semantic interoperability to enable searching across heterogeneous distributed databases. Inherently, there are limitations to crosswalks; there is rarely a one-to-one correspondence between the fields or data elements in different information systems."

-- Mary Woodley, *"Crosswalks: The Path to Universal Access?"*

26

Crosswalks

- Semantic mapping of elements between source and target metadata standards
- Metadata conversion specification: transformations required to convert metadata record content to another
  - Element to element mapping
  - Hierarchy and object resolution
  - Metadata content conversions
- Stylesheets are created to transform metadata based on crosswalks

27

Problems With Converted Records

- Differences in granularity (complex vs. simple scheme)
- Some data might be lost
- Differences in semantics
- Differences in use of content standards
- Properties may vary (e.g. repeatability)
- Converting may not always be the solution

28

Example: mapping MODS: title to DC: title

- Includes attribute for type of title
  - Abbreviated
  - Translated
  - Alternative
  - Uniform
- Other attributes: ID,authority,displayLabel,xLink
- Subelements: title, nameOfPart, numberOfPart, nonSort
- Title definition reused by: Subject, Related Item

29

Mapping mods: title to dc: title

- DC has one element refinement: alternative
- DC title has no substructure; MODS allows for subelements for partNumber, partName
- Best practice statement in DC-Lib says include initial article; MODS parses into <nonSort>
- MODS can link to a title in an authority file if desired

30

Metadata Crosswalks

- Dublin Core-MARC
- Dublin Core-MODS
- ONIX-MARC
- MODS-MARC
- EAD-MARC
- EAD-Dublin Core
- Etc.

31

Crosswalks

Library of Congress

http://www.loc.gov/marc/marcdocz.html

MIT

http://libraries.mit.edu/guides/subjects/metadata/mappings.html

Getty

http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html

32

II. MARC to Dublin Core Crosswalk (Unqualified).

Conventions: "$" is used to represent the control character subfield delimiter.

DC Element	MARC Fields	Implementation notes
Title	245	
Creator	100, 110, 111, 700, 710, 711	See Appendix 1 below; Contributor element not used.
	720	
Subject	600, 610, 611, 630, 650, 653	
Description	500-599, except 506, 530, 540, 546	
Contributor		See Appendix 1 below; Contributor element not used.
Publisher	260ab	
Date	260$c	
Type	Leader06, Leader07	See Appendix 2 for Leader-Type rules
	655	
Format	856$q	
Identifier	856$u	
Source	786ot	
Language	008/35-37	

III. MARC to Dublin Core Crosswalk (Qualified).

DC Element	DC Qualifier(s)	MARC Fields	Implementation notes
Title		245	
Title	Alternative	130, 210, 240, 242, 246, 730, 740	
Creator		100, 110, 111, 700, 710, 711	See Appendix 1 below.
		720	
Subject	LCSH	600, 610, 611, 630, 650	Second indicator=0
Subject	MeSH	600, 610, 611, 630, 650	Second indicator=2
Subject	LCC	050	
Subject	DDC	082	
Subject	UDC	080	
Description		500-599, except 505, 506, 520, 530, 540, 546	
Description	TableofContents	505	
Description	Abstract	520	First indicator=3
Contributor			See Appendix 1 below; Contributor element not used.
Publisher		260ab	

Metadata Good Practices

- Adherence to standards
- Planning for persistence and maintenance
- Documentation
  - Guidelines expressing community consensus
  - Specific practices and interpretation
  - Vocabulary usage
  - Application profiles
- Without good metadata and good practices, interoperability will not work

35

NISO's Metadata Principles

- **1**: Good metadata conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection.
- **2**: Good metadata supports interoperability.
- **3**: Good metadata uses authority control and content standards to describe objects and collocate related objects

36

NISO's Metadata Principles Continued

- **4**: Good metadata includes a clear statement of the conditions and terms of use for the digital object.
- **5**: Good metadata supports the long-term curation and preservation of objects in collections.
- **6**: Good metadata records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification.

37

Exercise

- Examine the MARC to DC crosswalk and the DC to MARC crosswalk, determining where data loss occurs

http://www.loc.gov/marc/marcdocz.html

38

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Dublin Core/MARC/GILS Crosswalk

Network Development and MARC Standards Office
Library of Congress

Date issued: 2001-03-12
Previous version: http://www.loc.gov/marc/dccross_199911.html

---

I. Introduction.

The following is a crosswalk between the fifteen elements in the Dublin Core Element Set and MARC 21 bibliographic data elements. In addition, it includes a crosswalk from Dublin Core to GILS attributes. The crosswalk may be used in conversion of metadata from another syntax into MARC. For conversion of MARC 21 into Dublin Core, many fields may be mapped into a single Dublin Core element. A MARC to Dublin Core Crosswalk is also available.

In the Dublin Core to MARC mapping, two mappings are provided, one for unqualified Dublin Core elements and the other for qualified. Qualifiers used are generally based on the qualifiers approved by the Dublin Core Metadata Initiative and documented in Dublin Core Qualifiers There are some qualifiers given (for Contributor, Creator, and Publisher) that have not been approved by the Dublin Core Metadata Initiative; as these are further standardized, this crosswalk will be adjusted.

MARC 21 fields are listed with field number, then two indicator values with field name/subfield name in parentheses. If both the field and subfield have the same name, the subfield name is not included. A blank (H'20') is indicated in this document by "#". The label is a shortened form of the element name. GILS attribute names for each Dublin Core element are also given. Note that the GILS mapping has not been revised since the April 1997 version of this document.

Definitions are taken from Dublin Core Metadata Element Set, Version 1.1: Reference Description. For further information about Dublin Core elements, including application notes (given in Comment), refer to that document. All Dublin Core elements are optional and repeatable. In this document elements are listed in alphabetical order by Dublin Core identifier (i.e. label).

II. Dublin Core to MARC and GILS Crosswalk.

Contributor

An entity responsible for making contributions to the content of the resource.

MARC 21:

Unqualified:

- 720 ##$a (Added Entry--Uncontrolled Name/Name) with $e=collaborator (or other term used as value of role qualifier)

Qualified:

- Personal: 700 1#$a (Added Entry--Personal Name) with $e=collaborator
- Corporate: 710 2#$a (Added Entry--Corporate Name) with $e=collaborator
- Conference: 711 2#$a (Added Entry--Conference Name) with $e=collaborator
- Role: 720 ##$e (Added Entry--Uncontrolled Name/Relator term
- Role (Personal): 700 1#$e (Added Entry--Personal Name/Relator term)
- Role (Corporate): 710 2$e (Added Entry--Corporate Name/Relator term)

Note: The above qualifiers have not been approved by DCMI.

GILS:

- Contributor

Coverage

The extent or scope of the content of the resource.

MARC 21:

Unqualified:

- 500$a (General note)

Qualified:

- Spatial: 522 ##$a (Geographic Coverage Note)
- Temporal: 513 ##$b (Type of Report and Period Covered Note/Period covered)

GILS:

- Supplemental Information
- Spatial: Bounding Coordinates
- Temporal: Time Period Textual

Creator

An entity primarily responsible for making the content of the resource.

MARC 21:

Unqualified:

- 720 ##$a (Added Entry--Uncontrolled Name/Name) with $e=author

Qualified:

- Personal: 700 1#$a (Added Entry--Personal Name) with $e=author
- Corporate: 710 2#$a (Added Entry--Corporate Name) with $e=author
- Conference: 711 2#$a (Added Entry--Conference Name) with $e=author
- Role: 720 ##$e (Added Entry--Uncontrolled Name/Relator term
- Role (Personal): 700 1#$e (Added Entry--Personal Name/Relator term)
- Role (Corporate): 710 2$e (Added Entry--Corporate Name/Relator term)

Note: The above qualifiers have not been approved by DCMI.

GILS:

- Originator

Date

A date associated with an event in the life cycle of the resource.

MARC 21:

Unqualified:

- 260 ##$c (Date of publication, distribution, etc.)

Qualified:

- Available: 307 ##$a (Hours, Etc.)
- Created: 260 ##$g (Date of manufacture)
- Issued: 260 ##$c (Date of publication, distribution, etc.)
- Modified: 583 ##$d with $a=modified
- Valid: 518 ##$a (Date/Time and Place of an Event Note). Text may be generated in $3 to include qualifier name.
- Scheme=ISO 8601: date may also be generated in 008/07-10; see below under Notes. If ISO 8601, use basic form that does not include hyphens in 008.

GILS:

- Date of Publication

Description
An account of the content of the resource.

MARC 21:

Unqualified:

- 520 ##$a (Summary, etc. note)

Qualified:

- Description.Abstract: 520 ##$a (Summary, etc. note)
- Description.TableofContents: 505 0#$a (Formatted Contents Note)

GILS:

- Abstract

Format
The physical or digital manifestion of the resource.

MARC 21:

Unqualified:

- 856 ##$q (Electronic Location and Access/Electronic format type)

Qualified:

- Extent: 300 ##$a (Physical Description)
  Note that "Extent" has been defined by the Format WG as "the size or duration of a resource"
- Medium: 340 ##$a (Physical Medium)
- Medium (Scheme=IMT): 856 ##$q (Electronic Location and Access/Electronic Format Type)

GILS:

- Available Linkage Type

Identifier

An unambiguous reference to the resource within a given context.

MARC 21:

Unqualified:

- 024 8#$a (Other Standard Identifier/Standard number or code)

Qualified:

- Scheme=URI: 856 40$u (Electronic Location and Access/Uniform Resource Locator)
- Scheme=ISBN: 020 ##$a (International Standard Book Number)
- Scheme=ISSN: 022 ##$a (International Standard Serial Number)
- Scheme=(other): 024 8#$a (Other Standard Identifier/Standard number or code) with $2=scheme value

Note: Only URI has been approved by DCMI.

GILS:

- Available Linkage

Language

A language of the intellectual content of the resource.

MARC 21:

Unqualified:

- 546 ##$a (Language note)

Qualified:

- Scheme=ISO 639-2: 041$a (Language code)
- Scheme=RFC 1766: 546 ##$a (Language note) with $b=RFC 1766

Note: RFC 1766 has been replaced by RFC 3066. When this change is reflected in DCMI Qualifiers, RFC 3066 should be used.

GILS:

- Language of Resource (note that GILS assumes use of Z39.53)

Publisher

An entity responsible for making the resource available.

MARC 21:

Unqualified:

- 260 ##$b (Publication, Distribution, etc. (Imprint)/Name of publisher, distributor, etc.)

Qualified:

- Personal: 700 1#$a (Added Entry--Personal Name) with $e=publisher
- Corporate: 710 2#$a (Added Entry--Corporate Name) with $e=publisher
- Conference: 711 2#$a (Added Entry--Conference Name) with $e=publisher

Note: The above qualifiers have not been approved by DCMI.
Note: It may be desirable to repeat a qualified publisher in 260$b

GILS:

- Distributor

Relation

A reference to a related resource.

MARC 21:

Unqualified:

- 787 0#$n (Nonspecific Relationship Entry/Note)

Qualified:

- Scheme=URI: 787 0#$o (Nonspecific Relationship Entry/Other identifier)
- HasFormat: 776 0#$n (Additional Physical Form Entry/Note)
- HasFormat: (Scheme=URI): 776 0#$o (Additional Physical Form Entry/Other identifier)
- IsFormatOf: 776 0#$n (Additional Physical Form Entry/Note)
- IsFormatOf: (Scheme=URI): 776 0#$o (Additional Physical Form Entry/Other identifier)
- IsPartOf: 773 0#$n (Host Item Entry/Note)
- IsPartOf (Scheme=URI): 773 0#$o (Host Item Entry/Other identifier)
- HasPart: 774 0#$n (Constituent Unit Entry/Note)
- HasPart (Scheme=URI): 774 0#$o (Constituent Unit Entry/Other identifier)
- IsVersionOf: 775 0#$n (Other Edition Entry/Note)
- IsVersionOf (Scheme=URI): 775 0#$o (Other Edition Entry/Other identifier)

- HasVersion: 775 0#$n (Other Edition Entry/Note)
- HasVersion (Scheme=URI): 775 0#$o (Other Edition Entry/Other identifier)
- IsBasedOn: 786 0#$n (Data Source Entry/Note)
- IsBasedOn (Scheme=URI): 786 0#$o (Data Source Entry/Other identifier)
- IsReferencedBy: 510 0#$a (Citation/References Note/Name of source)
- Requires: 538 ##$a (System Details Note)
- Replaces: 780 00$t (Preceding entry)
- IsReplacedBy: 785 00$t (Succeeding entry)

GILS:

- Cross Reference Relationship
- If scheme=URL: Cross Reference Linkage

Rights

Information about rights held in and over the resource.

MARC 21:

Unqualified:

- 540 ##$a (Terms Governing Use and Reproduction Note)

Qualified:

- Scheme=URL: 856 42$u (Electronic Location and Access/Uniform Resource Locator) with $3=Rights

GILS:

- Use Constraints

Source

A reference to a resource from which the present resource is derived.

MARC 21:

Unqualified:

- 786 0#$n (Data Source Entry/Note)

Qualified:

- Scheme=URI: 786 0#$o (Data Source Entry/Other identifier)

GILS:

- Sources of Data

Subject
The topic of the content of the resource.

MARC 21:

Unqualified:

- 653 ##$a (Index Term--Uncontrolled)

Qualified:

- Scheme=LCSH: 650 #0$a (Subject added entry--Topical term)
- Scheme=MeSH: 650 #2$a (Subject added entry--Topical term)
- Scheme=LCC: 050 ##$a (Library of Congress Call Number/Classification number)
- Scheme=DDC: 082 ##$a (Dewey Decimal Call Number/Classification number)
- Scheme=UDC: 080 ##$a (Universal Decimal Classification Number)
- Scheme=(other): 650 #7$a with $2=code from MARC Code List for Relators, Sources, Description Conventions

GILS:

- Uncontrolled Term

Title
The name given to the resource.

MARC 21:

Unqualified:

- 245 00$a (Title Statement/Title proper)
- If repeated, all titles after the first: 246 33$a (Varying Form of Title/Title proper)

Qualified:

- Alternative: 246 33$a (Varying Form of Title/Title proper)

GILS:

- Title

Type
The nature or genre of the content of the resource.

MARC 21:

Unqualified:

- 655 #7$a (Index Term--Genre/Form) with $2=local

Qualified:

- Scheme=DCMI Type: 655 #7$a (Index Term--Genre/Form) with $2=dct
- Scheme=(other): 655 #7$a (Index Term--Genre/Form) with $2=code from MARC Code List for Relators, Sources, Description Conventions

See Section III for use to determine Leader/06 (Type of Record) values.

GILS

- Medium

III. Notes.

In addition to the variable length fields listed in the mapping, a MARC 21 record will also include a Leader and field 008 (Fixed-Length Data Elements). Certain character positions in each of these fixed length fields of a USMARC record will need to be coded, although most will generate default values.

Leader: a fixed field comprising the first 24 character positions (00-23) of each record that provides information for the processing of the record. The following positions should be generated:

- Character Position 06: Type of record
  Leader/06 value should be set according to value in Type as follows (these values are from Dublin Core List of Resource Types (DC Type Vocabulary):

Type value	Leader/06 value
collection	p
dataset	m

event	r
image	k
interactive resource	m
service	m
software	m
sound	i
text	a

If no type is indicated, use value "a". If two type values are indicated, and one of these is "collection" use the other value for setting Leader/06. If more than two, use "m".

- Character Position 07: Bibliographic level
  - If Type value is collection, use value "c" (Collection)
  - All others, use value "m" (Monograph).
- Character Position 08: Type of control
  - Use value "#" (blank: no specific type of control).
- Character Position 09: Character coding scheme
  - Use value "#" (blank: MARC-8).
- Character Position 17: Encoding level
  - Use value "3" (Abbreviated level) or other value as appropriate to application
- Character Position 18: Descriptive cataloging form
  - Use value "u" (Unknown) to indicate that the descriptive cataloging form is unknown.

008 Fixed Length Data Elements: Forty character positions (00-39) containing positionally-defined data elements that provide coded information about the record as a whole or about special bibliographic aspects of the item being cataloged. For records originating as Dublin Core, the following character positions are used:

- Character positions 00-05: Date the MARC 21 record was created or converted (generate by date record entered system; formatted as YYMMDD)
- Character positions 07-10: Date of Publication (YYYY portion from Date if present). Qualified DC: Date.Issued in ISO 8601 (only YYYY portion).
- Character positions 35-37: Language. May be generated from data in Language if scheme=ISO 639-2.
- Other character positions can default to fill characters (ASCII 7C)

042$a Authentication Code: Use "dc" (identifies that MARC 21 record is derived from Dublin Core style record).

IV. Uses for mapping Dublin Core to MARC

A mapping between the elements in the Dublin Core and USMARC fields is necessary so that conversions between various syntaxes can occur accurately. Once Dublin Core style metadata is widely provided, it might interact with MARC records in various ways such as the following:

Enhancement of simple resource description record. A cataloging agency may wish to extract the metadata provided in Dublin Core style (presumably in HTML or XML) and convert the data elements to MARC fields, resulting in a skeletal record. That record might then be enhanced as needed to add additional information generally provided in the particular catalog. Some projects convert data and use as basic record for reporting to national bibliography.

Searching across syntaxes and databases. Libraries have large systems with valuable information in metadata records in MARC format. Over the past few years with the expansion of electronic resource over the Internet, other syntaxes have also been considered for providing metadata. It will be important for systems to be able to search metadata in different syntaxes and databases and have commonality in the definition and use of elements.

Go to:

- MARC Home Page
- Library of Congress Home Page

 **Library of Congress**
Library of Congress Help Desk (12/31/2002)

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

MARC to Dublin Core Crosswalk

February, 2001

Network Development and MARC Standards Office
Library of Congress

Table of Contents

I. Introduction.

The following is a crosswalk between core MARC 21 bibliographic data elements and elements in the Dublin Core Element Set. It may be used in conversion of metadata from MARC into Dublin Core. Since MARC is richer in data than Dublin Core, it differs from the Dublin Core/MARC Crosswalk in that multiple MARC fields are mapped to a Dublin Core element. The Dublin Core to MARC crosswalk maps a Dublin Core element to a single MARC field. In both crosswalks there are different mappings for Dublin Core simple or qualified. Not all possible MARC fields are included in this mapping, but only those considered useful for broad cross-domain resource discovery. Applications may wish to include other MARC elements that are prevalent in their data but are not listed here, or they may not include all that are listed.

MARC 21 fields are listed with field number with specific subfields if applicable. In many cases specific subfields are not provided, since applications may differ in subfields used. Applications may not need out control subfields such as $2, $5. Notes concerning implementation are given. Further information about the mapping is given at the end of the document. Definitions of MARC elements may be found at the MARC Bibliographic format site and definitions of Dublin Core elements from the Dublin Core Metadata Element Set Reference Description, Version 1.1 and Dublin Core Qualifiers. In this document MARC 21 elements are grouped together under the corresponding Dublin Core element in the order given at the Dublin Core reference description document. Part II gives a MARC to unqualified Dublin Core mapping, while Part III gives a mapping to qualified Dublin Core. Dublin Core qualifiers with no suitable mappings are not included; these may be reconsidered at a later date.

Note that it is not expected that round-trip mapping is possible using this crosswalk. Once MARC data is converted to Dublin Core, not enough information is retained to allow for mapping back to MARC accurately. This is inevitable when mapping a complex set of data elements to a simpler set. Where a tag is used in more than one place, a decision may need to be made which Dublin Core element to map it to.

II. MARC to Dublin Core Crosswalk (Unqualified).

Conventions: "$" is used to represent the control character subfield delimiter.

DC Element	MARC Fields	Implementation notes
Title	245	
Creator	100, 110, 111, 700, 710, 711	See Appendix 1 below; Contributor element not used.
	720	
Subject	600, 610, 611, 630, 650, 653	
Description	500-599, except 506, 530, 540, 546	
Contributor		See Appendix 1 below; Contributor element not used.
Publisher	260ab	
Date	260$c	
Type	Leader06, Leader07	See Appendix 2 for Leader-Type rules
	655	
Format	856$q	
Identifier	856$u	
Source	786ot	
Language	008/35-37	
	546	
Relation	530, 760-787ot	
Coverage	651	
	752	

Rights	506, 540	

III. MARC to Dublin Core Crosswalk (Qualified).

DC Element	DC Qualifier(s)	MARC Fields	Implementation notes
Title		245	
Title	Alternative	130, 210, 240, 242, 246, 730, 740	
Creator		100, 110, 111, 700, 710, 711	See Appendix 1 below.
		720	
Subject	LCSH	600, 610, 611, 630, 650	Second indicator=0
Subject	MeSH	600, 610, 611, 630, 650	Second indicator=2
Subject	LCC	050	
Subject	DDC	082	
Subject	UDC	080	
Description		500-599, except 505, 506, 520, 530, 540, 546	
Description	TableofContents	505	
Description	Abstract	520	First indicator=3
Contributor			See Appendix 1 below; Contributor element not used.
Publisher		260ab	
Date	Created	260cg	
		533$d	
Date	Issued	260$c	
		008/07-10	
Type	DCMI Type Vocabulary	Leader06, Leader07	See Appendix 2 for Leader-Type rules
		655	Subfield $2=dct

Format	IMT	856$q	
	Extent	300$a	
		533$e	
	Medium	340$a	
Identifier	URI	856$u	
Source	URI	786$o	
Language	ISO 639-2	008/35-37	
		041	Multiple codes need to be parsed by threes.
	RFC1766	546	
Relation	IsVersionOf	775,786nt	
Relation	IsVersionOf URI	775,786$o	
Relation	Has Version	775nt	
Relation	HasVersion URI	775$o	
Relation	IsReplacedBy	785nt	
Relation	IsReplacedBy URI	785$o	
Relation	Replaces	780nt	
Relation	Replaces URI	780$o	
Relation	Requires	538	
Relation	IsPartOf	760,773nt	
		440, 490,800,810,811,830	
Relation	IsPartOf URI	760,773$o	
Relation	HasPart	774nt	
Relation	HasPart URI	774$o	
Relation	IsReferencedBy	510	

Relation	IsFormatOf	776nt	
Relation	IsFormatOf	530	
Relation	IsFormatOf URI	776$o	
		530$u	
Relation	HasFormat	776nt	
		530	
Relation	HasFormat URI	776$o	
		530$u	
Coverage	Spatial	522, 651	
		255	Some 255 information equivalent to DC encoding scheme but different syntax
		650$z	
		752	
Coverage	Spatial ISO 3166	043$c,044$c	Defined in MARC in January 2001.
	Spatial TGN	651	Subfield $2=tgn
Coverage	Temporal	513$b	
		033$a	
Rights		506, 540	No qualifiers defined.

Appendix 1--Notes

Creator vs. Contributor. In this mapping MARC name fields are mapped to DC Creator rather than Contributor for the following reasons:

- Creator elements are used for all the MARC fields indicated, both 1XX and 7XX.
- It is possible that in the future a distinction might be made between Creator and Contributor on the basis of the value of a role element. However, DCMI has not defined any qualifiers for specifying role as of the date of this mapping.
- Creator has been chosen, since it is more likely that the name included would have made a primary contribution rather than a secondary one.

Identifier. DCMI has only defined an encoding scheme for URI for this element.

Element encoding schemes. Some encoding schemes defined by DCMI have not been included because they are not generally found in MARC data. Examples include DCMI Point, DCMI Box.

Appendix 2 - Conversion rules for Leader06 - dc:Type mapping.

Multiple Type fields may be used; conversions below may result in 2 or 3 Type fields. There are several additional sources of type information in the MARC record; only coded values in Leader/06 and Leaader/07 are detailed in this chart. Field 655 may also be used for more specific type information.

Leader/06 value	Leader/07 value	Type value
a,c,d,t		text
e,f,g,k		image
i,j		sound
m,o,p,r		no type provided
p	c,s	collection

Institutions may want to consider generating additional type values, such as "map" or "cartographic" for codes e or g; "musical notation" for codes c or d, etc.

Go to:

- Dublin Core to MARC Crosswalk
- MARC Home Page
- Library of Congress Home Page

Library of Congress

Library of Congress Help Desk (04/12/2007)

Terminology

Metadata: In general, data about data; functionally, structured data about data. Metadata includes data associated with either an information system or an information object for purposes of description, administration, legal requirements, technical functionality, use and usage, and preservation. Traditional library cataloging is a form of metadata. (Dublin Core Metadata Initiative glossary)

Schema: Defines the vocabulary of a particular set of metadata (i.e., element names and formatting rules). A schema is usually defined by some authority to describe data in a standard way so that it may be accessed by other users or applications. (Tom Sheldon's Linktionary.com)

Metadata Schema: Sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. The definition or meaning of the elements themselves is known as the semantics of the scheme. The value given to the metadata elements are the content. (Understanding Metadata)

Dublin Core Metadata Element Set: Includes terms intended to facilitate the discovery of resources. The metadata is associated with the intellectual content, intellectual property, and/or instantiation characteristics of an information resource. (Dublin Core Metadata Initiative glossary)

Standard Generalized Markup Language (SGML): A non-proprietary language/enabling technology for describing information. Information in SGML is structured like a database, supporting rendering in and conversion between different formats. Both XML and later versions of HTML are instances of SGML. (Dublin Core Metadata Initiative glossary)

Hypertext Markup Language (HTML): The standard text-formatting language for documents on the World Wide Web. HTML text files contain content that is rendered on a computer screen and markup, or tags, that can be used to tell the computer how to format that content. HTML tags can also be used to encode metadata and to tell the computer how to respond to certain user actions, such as a mouse click. (Dublin Core Metadata Initiative glossary)

eXtensible Markup Language (XML): A subset of SGML. XML is designed to bring the power and flexibility of generic SGML to the World Wide Web, while maintaining interoperability with full SGML and HTML. (Dublin Core Metadata Initiative glossary)

XML Schemas: Express shared vocabularies and allow machines to carry out rules made by people. They provide a means for defining the structure, content and semantics of XML documents. (World Wide Web Consortium Web page)

Namespaces *An effort to allow markup from different XML applications to be used in the same document without conflict (even if element names used in each namespace are the same).*

XSL *Extensible Stylesheet Language, an XML application for transforming XML documents into a form that could be viewed in a web browser. Eventually split to form: XSLT and XSL-FO.*

XSL-FO *An XML application for describing the layout of both printed pages and webpages, sometimes compared to PostScript.*

CSS *Cascading Style Sheets, used originally for HTML, and, when XML was invented, it was used for that, too.*

XLink *Extensible Linking Language, used to define more powerful linking constructs to connect XML documents, going beyond the "a" tag used in HTML.*

XPointer *Addresses individual parts of an XML document.*

XPath *Extracted from the addressing parts of XLink and XPointer.*

Metadata Encoding and Transmission Standard (METS): XML schema for encoding descriptive, administrative, and structural metadata regarding objects within a digital library. (METS Web page)

Metadata Object Description Schema (MODS): XML schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. MODS is intended to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than numeric ones, in some cases regrouping elements from the MARC 21 bibliographic format. (MODS Web page)

Metadata Authority Description Schema (MADS): XML schema for an authority element set that may be used to provide metadata about agents (people, organizations), events, and terms (topics, geographics, genres, etc.). MADS was created to serve as a companion to the Metadata Object Description Schema (MODS). As such, MADS has a relationship to the MARC 21 Authority format, as MODS has to MARC 21 Bibliographic. Both carry selected data from MARC 21**.** (MADS Web page)

Web Style Sheets: Describe how documents are presented on screens, in print, or perhaps how they are pronounced. By attaching style sheets to structured documents on the Web (e.g. HTML), authors and readers can influence the presentation of documents without sacrificing device-independence or adding new HTML tags. (World Wide Web Consortium Web page)

XSLT: Language for transforming XML documents into other XML documents. A transformation expressed in XSLT is called a style sheet. (World Wide Web Consortium Web

page)

MARC XML: Framework for working with MARC data in an XML environment. This framework is intended to be flexible and extensible to allow users to work with MARC data in ways specific to their needs. The framework will contain many components such as schemas, style sheets, and software tools. (MARC XML Web page)

Validation Checking a document against a schema or DTD.

Well-formed A document is considered 'well-formed' if it satisfies XML grammatical rules, including where tags are placed (especially beginning and ending), what the tags look like, what element names are legal, etc.

Parser A program that divides the XML document up into individual elements, attributes, etc., to determine if well-formed.

Validating parser Checks for both well-formed XML and conformance with a schema.

DTD Document Type Definition, used first with SGML. Also used with XML but gradually being phased out in favor of schemas.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

References Handout: Course 2

Louis Rosenfeld and Peter Morville, Information Architecture for the World Wide Web: Designing Large-Scale Web Sites. Cambridge, Mass.: O'Reilly, 2002.

Peter Morville, Ambient Findability: What We Find Changes Who We Become. Cambridge, Mass.: O'Reilly, 2005.

Godfrey Rust and Mark Bide, The <INDECS> Metadata Framework: Principles, Model and Data Dictionary. Available at: http://www.doi.org/topics/indecs/indecs_framework_2000.pdf

NISO. Understanding Metadata. http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

Tillett, Barbara B. *A taxonomy of bibliographic relationships*. Library Resources & Technical Services, 35(2):150--158, 1991.

IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional Requirements for Bibliographic Records: Final report. Available at: http://www.ifla.org/VII/s13/frbr/frbr.pdf

DCMI Abstract Model. http://dublincore.org/documents/abstract-model/

MARC 21. Available at: http://lcweb.loc.gov/marc/

Dublin Core. Available at: http://dublincore.org/

Metadata Object Description Schema (MODS). Available at: http://www.loc.gov/standards/mods/

IEEE 1484 Learning Objects Metadata (IEEE LOM). Official standard available at: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

Simpler chart and DC mapping available at: http://www.ischool.washington.edu/sasutton/IEEE1484.html

CanCore Guidelines (an AP of IEEE-LOM, best guidelines available for this schema). Available at: http://www.cancore.ca/en/

ONIX for Books. Available at: http://www.editeur.org/onix.html

XML: Extensible Markup Language. Available at: http://www.w3.org/XML/

XML Tutorial. Available at: http://www.w3schools.com/xml/default.asp

XHTML: Extensible Hypertext Markup Language. Available at: http://www.w3.org/TR/xhtml1/

METS: Metadata Encoding & Transmission Standard. Available at: http://www.loc.gov/standards/mets/

Open Archives Initiative http://openarchives.org

The OpenURL Framework for Context-Sensitive Services http://www.niso.org/committees/committee_ax.html

Metadata Mappings (Crosswalks) http://libraries.mit.edu/guides/subjects/metadata/mappings.html

Metadata: Mapping Between Metadata Formats http://www.ukoln.ac.uk/metadata/interoperability/

MARC 21 to Dublin Core Crosswalk http://www.loc.gov/marc/marc2dc.html

Dublin Core/MARC/GILS Crosswalk http://www.loc.gov/marc/dccross.html

Margaret St. Pierre and William P. La Plant, *Issues in crosswalking content metadata standards*. (NISO White Paper). Bethesda, Md.: National Information Standards Organisation, 15 October 1998.
Available from: http://www.niso.org/press/whitepapers/crsswalk.html

Carol Jean Godby, Devon Smith, and Eric Childress, *Two Paths to Interoperable Metadata*, a paper presented at the 2003 Dublin Core Conference http://www.oclc.org/research/publications/archive/2003/godby-dc2003.pdf

ANSI/NISO Z39.19 - 2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies http://www.niso.org/standards/resources/Z39-19.html

SKOS Core Guide http://www.w3.org/TR/swbp-skos-core-guide/

SKOS Quick Guide to Publishing a Vocabulary on the Semantic Web http://www.w3.org/TR/swbp-thesaurus-pubguide/

DCMI Registry http://dublincore.org/groups/registry/

CEN Workshop Agreement 14855: Dublin Core Application Profile Guidelines. http://www.cenorm.be/cenorm/businessdomains/businessdomains/iss/cwa/cwa14855.asp

CEN Workshop Agreement 15248: Guidelines for machine processable

representation of Dublin Core Application Profiles.
http://www.cenorm.be/cenorm/businessdomains/businessdomains/isss/cwa/cwa+15248.asp

Thomas Baker and Makx Dekkers, Identifying metadata elements with URIs: the CORES Resolution. D-Lib magazine, July/August 2003.
http://www.dlib.org/dlib/july03/baker/07baker.html

Guidelines for Assigning Identifiers to Metadata Terms
http://www.ukoln.ac.uk/metadata/dcmi/term-identifier-guidelines/

DCMI Mixing and Matching FAQ
http://www.ukoln.ac.uk/metadata/dcmi/mixing-matching-faq/

DLF OAI Best Practices Guidelines http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl?PublicTOC

Using Dublin Core http://dublincore.org/documents/usageguide/

MODS User Guidelines, vers. 3 http://www.loc.gov/standards/mods/v3/mods-userguide.html

DLF MODS Implementation Guidelines For Cultural Heritage Materials
http://www.diglib.org/aquifer/DLF_MODS_ImpGuidelines_ver4.pdf

Analyzing Metadata for Effective Use and Re-Use, by Naomi Dushay, Diane I. Hillmann. Available at: http://dc2003.ischool.washington.edu/Archive-03/03dushay.pdf

Spotfire [software] Available at: http://www.spotfire.com/

National Science Digital Library Metadata Guidelines
http://metamanagement.comm.nsdl.org/outline.html

DLESE Metadata page http://www.dlese.org/Metadata/

RDA developments : http://www.collectionscanada.gc.ca/jsc/index.html

Readings in metadata and cataloging education:
http://www.loc.gov/catworkshop/readings.html

Rebecca Guenther, Library of Congress: Using Metadata Standards in Digital Libraries: Introduction to METS, MODS, PREMIS and MIX :
http://www.loc.gov/standards/mods/presentations/litaprogram-an2007.html

Working Group on the Future of Bibliographic Control:
http://www.loc.gov/bibliographic-future/

PREMIS: http://www.loc.gov/standards/premis/

Metadata Standards and Applications

Handout: Monitoring and Participating in Metadata Developments

Recommended e-journals, current awareness resources and discussion lists

1. D-Lib Magazine (http://www.dlib.org/)
2. Ariadne (http://www.ariadne.ac.uk/)
3. Current Cites (http://lists.webjunction.org/currentcites/) Also available as mailing list or RSS feed
4. Journal of Digital Information (http://jodi.tamu.edu/)
5. NISO Newsline (http://www.niso.org/news/newsline/)
6. MetadataLibrarians listserv (http://metadatalibrarians.monarchos.com/)

Blogs

Sign on with a blog aggregator (Bloglines is a good basic one) and start reading. Some recommendations:

1. Lorcan Dempsey's weblog: (http://orweblog.oclc.org/)
2. Inquiring Librarian: (http://inquiringlibrarian.blogspot.com/)
3. Weibel Lines: (http://weibel-lines.typepad.com/weibelines/)
4. The Shifted Librarian (http://www.theshiftedlibrarian.com/)

Recommended Format Specific Discussion Lists

1. DC-General (consider a Working Group list as well) (http://dublincore.org/about/contact/#dcgeneral)
2. MODS list (http://listserv.loc.gov/listarch/mods.html)
3. VRA Core (http://vraweb.org/vra-l/index.html)

Recommended Conferences

1. ALA and the specialized library divisions (especially LITA and ALCTS)
2. Dublin Core (next is DC 2008: http://dc2008.de/
3. Digital Library Federation forums (http://www.diglib.org/forums.htm)

Organize a local forum where you and your peers can have regular discussions about the work you're doing (e.g., Cornell Metadata Working Group http://metadata-wg.mannlib.cornell.edu/)