Cataloging for the 21st Century -- Course 2

Metadata Standards & Applications

Instructor Manual

Developed by the
Association of Library Collections & Technical Services
and the
Library of Congress

This Catalogers Learning Workshop (CLW) version is the official Cataloging for the 21st Century (Cat21) version of this course

Library of Congress
Catalogers Learning Workshop
Washington, DC
February 2009

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
9:00-9:30	**Registration. Breakfast**
9:30-9:45	**Getting started; Introductions, Orientation and Background**
9:45-10:30	**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
10:30-10:45	Break
10:45-12:30	**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.
12:30-1:30	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
9:30-11:30 (with break 10:30-10:45)	**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 Exercise: Analyze descriptive metadata application profiles
11:30-12:15	**6. Controlled Vocabularies** • 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

12:15-1:15	Lunch
1:15-2:30	**7. Approaches to Metadata Creation, Storage and Retrieval** • 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
2:45-4:30	**8. Metadata Interoperability and Quality Issues** • Understand interoperability protocols (OAI-PMH for harvesting, OpenURL for references) • Introduction to crosswalking and mapping • Discuss the criteria that can be used to determine quality in metadata Exercise: Evaluate a small set of human and machine-created metadata **OR** Examine the MARC to DC crosswalk and DC to MARC crosswalk and discuss where metadata loss occurs.

Introduction and Notes for Instructors

Introduction

Metadata Standards and Applications has been designed for those with a background in traditional libraries (particularly, but not solely, cataloging) who need additional background in metadata and digital libraries. An optimum class size of 25-35 is suggested, with two trainers preferred (at least one should have some strong experience with technical applications like XML and OAI). Two days is preferred to address all the material with sufficient time for exercises, but it is possible to shorten the class to a day and a half by shortening the specific metadata sessions and/or shortening or omitting the vocabulary portion. It is virtually impossible to do this course without Internet access for the instructor.

The order of presentation is not necessarily cast in stone, and instructors should feel free to change the order of presentation to suit their needs. The introductory material can be speeded up, depending on the audience and its familiarity with how websites are constructed. The exercises in parts 1 & 7 can be shortened, rearranged or different sites used if instructors desire.

There are certain assumptions inherent in this course, the most important being that the students have some experience with the traditional "culture" of libraries, and the way traditional libraries organize and present their materials to users. Those without some exposure to library cataloging practices will be at a distinct disadvantage in this course without some preparation. Two good resources to prepare anyone for this course would be:

1. *Understanding Metadata* (http://www.niso.org/standards/resources/UnderstandingMetadata.pdf). It is short, introduces some relevant standards, and also has a good reading list, and a glossary.
2. Arlene Taylor's *The Organization of Information*, 2nd ed. (Libraries Unlimited, 2004), particularly Chapters 4 ("Encoding Standards"), 6 ("Metadata"), and 7 ("Metadata: Description").

As is generally the case, those who "know what they don't know" and have a good background in traditional libraries should find most of this course accessible. The course tries to maintain a balance between abstract concepts and experiential understanding, with some attention to technical concepts that have likely been encountered before but incompletely understood. Throughout there is strong attention to relating what attendees already know about traditional library practice to what they desire to know more about in the metadata world. Oftentimes this is done by approaching what they do know in a more abstract manner than they are used to (MARC is a good example) and showing the similarities and differences to other metadata schemas they need to know more about.

Because there is generally not much of an opportunity for “hands on” work in this course, it is not sufficient for those who will need to create metadata immediately upon their return to their jobs, but it should allow them to incorporate the learning necessary to do that fairly quickly, and certainly with more confidence. In that sense, it is (and is not) what we might term a “basic course.” It assumes a basis for learning new concepts, but does not (and cannot) take attendees directly into metadata practice. The References document and the Ongoing Learning document provide suggestions for how to move on from the course materials towards increased knowledge.

Preparing to Teach this Course

Even for those well prepared to teach the material, this course requires a fair amount of preparation to do well. Instructors should study the exercise answer sheets before teaching the class. The schedule provided is only a sample, since times may differ.

Other preparation time consists of exploring the sites chosen for discussion in Sections 1, 5 & 7 to the point where the notes can be used smoothly (or modified). The instructor should familiarize him/herself with the preservation metadata examples in session 3, as this may be new material. Preparation is also recommended for the OAI demonstration, although slides are included that can take the place of a live demonstration if necessary or preferred. If a live demo is done, those slides can be eliminated from the presentation, though they are still helpful for the handout.

It is helpful to do an assessment of student needs before the class. Please make sure the class fills out the evaluation provided at the back of the manual.

Detailed Notes on Course Components

The notes below are general in nature and focus particularly on the exercises. Additional notes are included with slides.

Session 1: This section can be managed in a short period of time, leaving more time for the more challenging material coming later.

Exercise: Examine three digital library instances, list differences in user approach and experience and contrast with student’s home library

Note: This and the exercise for Section 2 works best as a full group exercise, lead by the instructor. Below are notes to assist in leading the exercises. Any search suggestions can be substituted with others of more interest to the instructor.

1. Alsos: Digital Library for Nuclear Issues **(**http://alsos.wlu.edu/default.aspx)

Note: This site is distinctive because it provides combined access to both digital and physical resources (books, primarily)

Suggestion: From the front page, note the various entry paths into the collection: search, browse, topics of the month, and the overviews. Select one of the browse topics (ex.: People) and navigate to the entry for Bethe, Hans, a noted physicist. From the available resources, choose the book on Robert Oppenheimer. Note the evidence of metadata both in the description of the book, and the lists of topics on the right (this explains why the book showed up under Hans Bethe even though it was about Robert Oppenheimer). Point out that this is a more granular level of description than one would expect from traditional library cataloging.

2. CSUN Oviatt Library: Digital Collections (http://library.csun.edu/Collections/SCA/digicoll.html)

Note: This site illustrates a common library-based model for exposing digital collections.

Suggestion: Click on the first collection (San Fernando Valley History). Note two sets of upper level categories: Natural resources, social life and customs, etc. & Topics, communities, people, etc. Click on one of the first group of topics (ex. Transportation), note that there is a search box here, just as on the collection page, but that the instructions suggest cutting and pasting terms (this is probably a structured subject based search, while the first page may be keyword). Question: Why aren't the terms linked to provide canned searches?

Search on a term (ex.: Airports), and note the navigational elements above and below the thumbnails. Click on a linked title below the thumbnails. This will take you to a larger image and metadata for the photograph. Note the distribution of linked and unlinked terms, and discuss what might be the reasoning behind each choice.

3. Birdsource (http://www.birdsource.org/)

Note: This site is very data driven, and supports far more interaction than either of the others.

Suggestion: Click on e-Bird, note the navigation bar on top and the links down the right side. Note the variety of different users supported, from beginners looking for descriptions of birds, to advanced birders and researchers looking for data. Choose "view and explore data" from the top navigation bar, and note the choices. Choose the view species option, and type in a bird species (ex.: indigo bunting). Warning: it will take a few seconds for the data to load. Click on the link to change location, and choose a state or smaller entity.

Suggestion: Return to e-Bird home, and select "All About Birds" from the right hand menu and select "Bird Guide." You'll be directed to a search page which offers a choice of pull down lists: taxonomic order, and alphabetic. Choose alphabetic and select a bird (ex.: Indigo Bunting). Note the two approaches: summary and detailed, and the sound, image and text data available.

For sessions 2, 3, and 4 there are 2 objects that are the focus of the exercise. For each session it is best if the instructor walks through one of the examples and gives the other for the exercise. The students may work in groups if they prefer. Templates are provided. The student manual includes a printout of the object with instructions on how to do the exercise and a template to fill in.

The instructors' and students' manual includes the answer sheets. They are located in the back of the manual after the other materials. It may be better to inform the students that the sheets are there after the Session 2 exercise, and encourage them not to look ahead at the answers for the upcoming sessions.

Session 2: Exercise. In session 2 trainees will describe one in either MODS or Dublin Core. After walking through the answers to one of the examples, the instructor could divide the group with half doing MODS and half doing Dublin Core. Instructors may substitute different objects instead of what is provided. Instructors may also ask some in the group to choose a different standard for description (e.g. VRA) and provide group with a list of data elements in that standard. A template is provided to fill in the element name and value.

Session 3: Exercise. The same examples are used for PREMIS metadata. The instruction sheet for trainees for this example includes the list of PREMIS data elements to be used and gives information about the objects which gives some of the values to fill in. The information about relationships is needed to determine how the files relate to each other. The same template is used to fill in element name and value. This may be difficult for students if they are encountering technical metadata for the first time; the instructor should at least walk through one example and may wish to just go over the answers for both with the students rather than have them complete them independently.

Session 4: Exercise. This is for trainees to put the metadata supplied above into XML coding. Again, divide up the group between MODS and Dublin Core. There is a DC and MODS template with XML element tags and space to fill in the appropriate values.

For the above exercises, if possible it is helpful to put up the object on the screen for trainees to look at as they prepare the answers.

Other exercises are discussed in class and included in the slides. Instructors' may substitute other exercises as desired.

Session 5: In the last part of the session, there is a case study of how an application profile is used to drive presentation and display of digitized material. It is important that the instructor read the instructors' notes in the powerpoint slides to be able to explain how the metadata pieces fit together.

Exercise. Trainees will analyze some example descriptive metadata application profiles. Give the students time to look at these (they are included in the packet, although only part of U Maryland's is because of its length). Divide the class into 3 groups and each group will analyze one profile. Alternatively this may be done as a group. See the questions to address on the last slide in this session. Points for discussion may include what works and what doesn't work in each profile, how thorough is the documentation, is it easy to apply, are there minimal requirements, are there examples.

Session 6: No exercise.

Session 7: Exercise: From example websites, determine the underlying models for each. Please feel free to select sites that are more familiar to you.

1. Plant and Insect Parasitic Nematodes: http://nematode.unl.edu/

Note: A fairly traditionally organized, HTML based site. Choose "View Source" on the front page to see the table based front page navigation. There is no search capability, and most likely no metadata.

2. Public Radio Market: http://www.prms.org/

This is a fairly heavily metadata driven site. Search is separated into classical and other genres (get the group to discuss why that might be, and perhaps do some sample searches). Note the top and side menus and features, particularly the tie-ins with NPR. Try a search (suggestion: Cape Verde), and point out that the site requires you to designate where your support percentage will go (it saves time to invoke this site prior to presenting the material, and go through the initial pages prior to showing it to the class).

From the Cape Verde search results, click on the first album, note that there are sound files listed for all the tracks, plus a tab for credits. Discuss with the group how this granularity level differs from traditional library cataloging, and where the metadata comes from.

3. Alcohol, Temperance and Prohibition (Brown University Library Center for Digital Initiatives: http://dl.lib.brown.edu/temperance/

This is a fairly common approach to presenting and providing access to digitized collections using an XML based approach.There are rich metadata records behind the resources. This gives an option for simple search and advanced search. The advanced one allows for applying limits to the search. Consider the ability to access the items in the collection through the search and browse functions. Try searching "beer" in simple search and see how the results are presented with metadata and a thumbnail of the image. Then try advanced search with "beer" and limit by years 1920-1930.

Section 8:

OpenURL demo: there are screenshots of a demo using the MARCEdit tool. If desired the instructor could do a live demo. See instructors' notes for the OAIster demo.
* Exercise_sect.8.doc
* Webpages:
    http://www.nbtc.cornell.edu/mainstreetscience/grab_and_go/treeswinterpg.htm
    http://eg2.ischool.washington.edu/registry
    http://ithacasciencezone.com/

Exercise: Evaluate a small set of human and machine-created metadata. The prepared two page handout should be augmented with the printouts of the associated Webpages. (as of this Aug. 2008 revision, the websites associated with these were no longer available—there may be replaced examples in a revised version). This exercise may be omitted if desired, replaced with other examples, or replaced with the alternative exercise (see below).

Metadata Quality exercise:

Example 1: The first metadata record was created by the iVia service, an automated metadata creation tool. Note that the title was the html title for the page, which was reused for each of the resources in the collection and so has the same title as all the other pages in the collection. The iVia service creates LCSH, and also "key phrases" which are all lumped together under Subject in this display (though they're in separate instances in the record itself). Note that the descriptive paragraph was taken from the actual webpage. The "HTML Title" noted here is actually the collection title.

Example 2: This record for the NSDL Registry was created by another automated tool: dc-dot. This example uses XHTML (the tool can output in a number of syntaxes). Note that the record uses both DC namespaces so is qualified DC (primarily because it identifies some syntax and vocabulary encoding schemes—W3CDTF, DCMIType, URI. Note also that the subject terms are culled from the page and include a number of navigation headings as well as personal names and significant words.

Example 3: Note that this record is exactly as it appeared on the web. The origin of this record is unknown, although it bears signs of having been crosswalked from a database of some kind. It is in no recognized schema, and carries several forms of name expressions. Note the value under Media (those older than about 45 will recognize that this is likely a 33 1/3 LP format).

Example 4: This record for the ICSD Sciencezone is a human created DC collection record corresponding to a website of materials. Note that it is quite spare, although it includes information not likely to have been created by machine (Grade Level, which is actually educationLevel in DC). Also note that the labels in the yellow column only roughly correspond to DC elements. "HTML Title" is the same as Title because NSDL recently started using that for collection titles, but then repeats them for some reason for actual collection records.

Alternative Exercise: Examine the MARC to DC crosswalk and DC to MARC crosswalk and discuss where metadata loss occurs. These crosswalks are included in the packets.

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

Exercise: Analyze descriptive metadata application profiles

6. Controlled Vocabularies

- 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

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

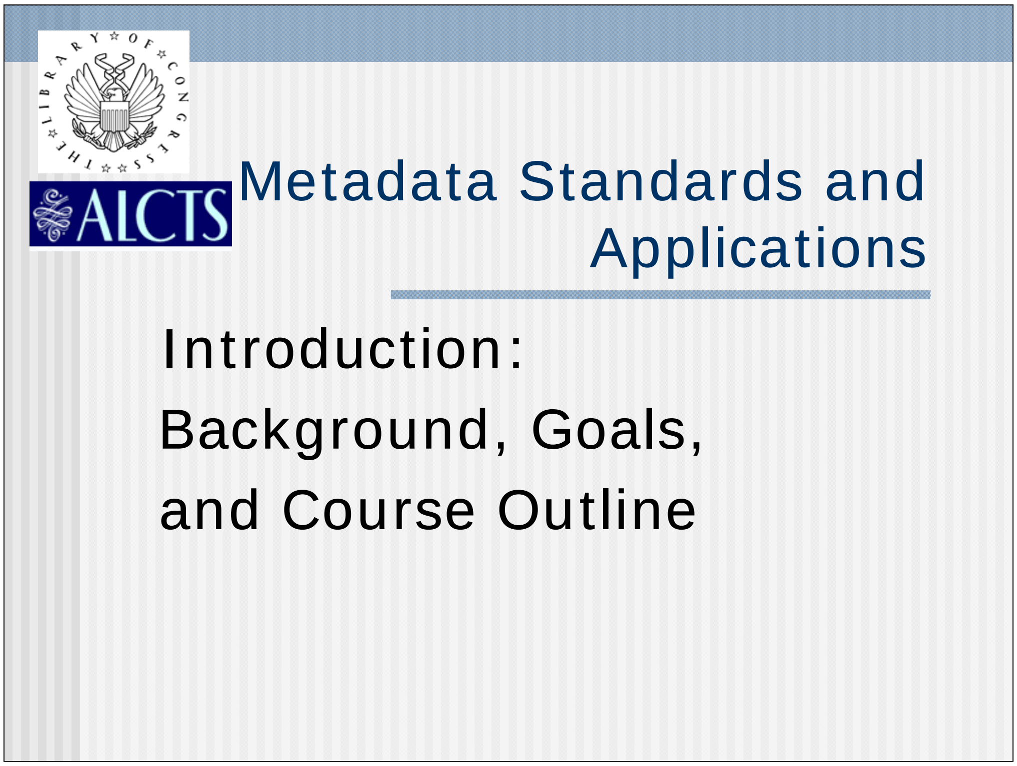
Exercise: Using example websites, determine the underlying models for each

8. Metadata Interoperability and Quality Issues

- Understand interoperability protocols (OAI-PMH for harvesting, OpenURLs for references)
- Introduction to crosswalking and mapping
- Discuss the criteria that can be used to determine quality in metadata

Exercise: Evaluate a small set of human and machine-created metadata **OR** Examine the MARC to DC crosswalk and DC to MARC crosswalk and discuss where metadata loss occurs.

THIS PAGE INTENTIONALLY LEFT BLANK FOR DOUBLE SIDED COPY



Suggested activities for getting started, before beginning slide content:

•Introduce instructors.

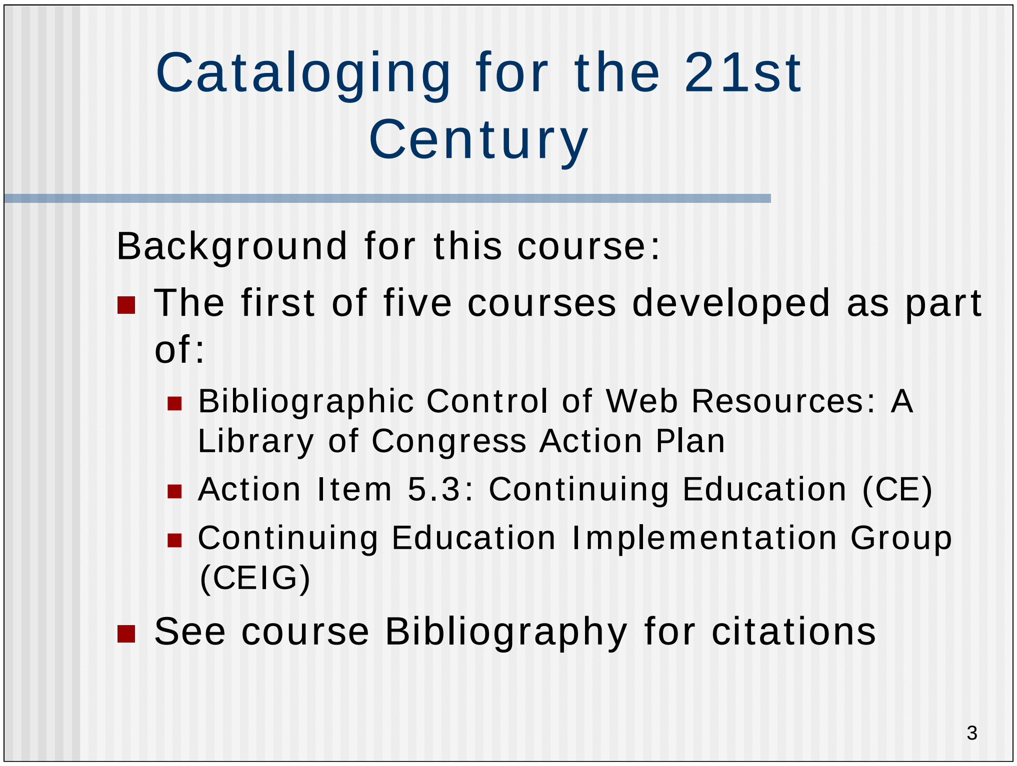
•Ask each participant to introduce him/herself, identifying his/her institution and position, and also, why they are there—saying a little about the connection between their job and interest in this course.

•Go over with participants the contents of their manuals.

  •Point out the table of contents, and the presence of the examples, exercises, glossary, bibliography.

Version 2.1, February 2009

2


Cataloging for the 21st Century
Background for this course:
The first of five courses developed 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
3

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

- 1. Rules and Tools for Cataloging Internet resources
- 2. Metadata Standards and Applications
- 3. Principles of Controlled Vocabulary and Thesaurus Design
- 4. Metadata and Digital Library Development
- 5. Digital Project Planning and Management Basics

4

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

5

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 quality issues

6

Digital library design and development is a process, rather than a recipe that can be followed.

Within that process, this workshop explores the role of metadata, and the roles and responsibilities of the metadata specialist.

Situate this course compared to:

Course 4: Metadata and Digital Library Development

Course 5: Digital Project Planning and Management Basics

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

7

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
  - Specific descriptive metadata formats
  - Relationship models

8

Outline of this course cont.

- Session 3. Technical and Administrative Metadata Standards
- Session 4. Metadata Syntaxes and Containers
- Session 5. Application Profiles and how they are used in digital libraries

9

Outline of this course cont.

- Session 6. Controlled Vocabularies
- Session 7. Metadata Creation, Storage and Retrieval
- Session 8. Metadata Interoperability and Quality Issues

10

1. Introduction to Digital Libraries and Metadata

Metadata Standards and Applications Workshop

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

We are all quite familiar with the traditional library; brick and mortar (buildings) with books and information items of various formats.

But, what are our thoughts concerning digital libraries or as some may say “electronic libraries” or “virtual libraries?

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

Definitions from Wikipedia

A better definition could be…[as read above]

This is DLF's "working" definition of a digital library.

DLF continues to state that " the concept of digital library has multiple senses that one might invoke in various contexts. For example, the concept may refer simply to the notion of collection without reference to organization, intellectual accessibility or service attributes. This extended sense seems to be in play, for example, when we hear the World Wide Web described as a digital library. The concept might also refer to the organization underlying the collection, or even more specifically to the computer-based system in which the collection resides."

DLF is a <u>membership</u> consortium of libraries and related agencies that are pioneering the use of electronic information technologies to extend collections and services. Your institution must join to participate in activities.

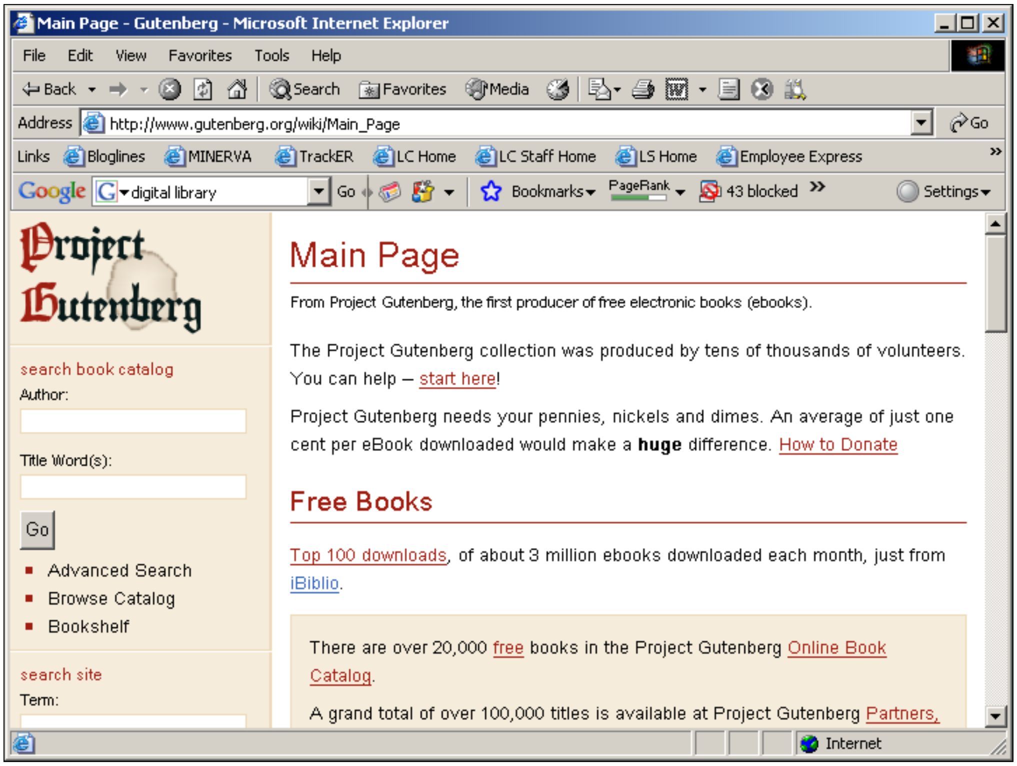
The digital library is not a single entity;

•The digital library requires technology to link the resources of many

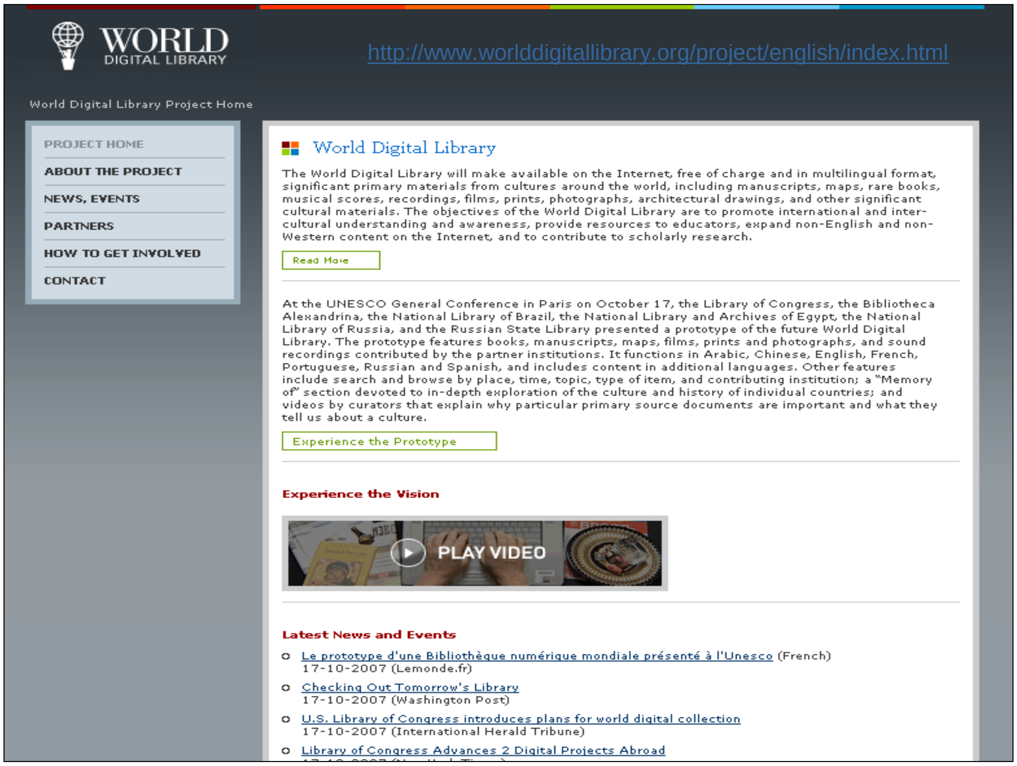
* The linkages between the many digital libraries and information services are transparent to the end users;

•Universal access to digital libraries and information services is a goal;

* Digital library collections are not limited to document surrogates: they extend to digital artifacts that cannot be represented or distributed in printed formats



Wikipedia even claims that Project Gutenberg is the first digital library, established in 1971.



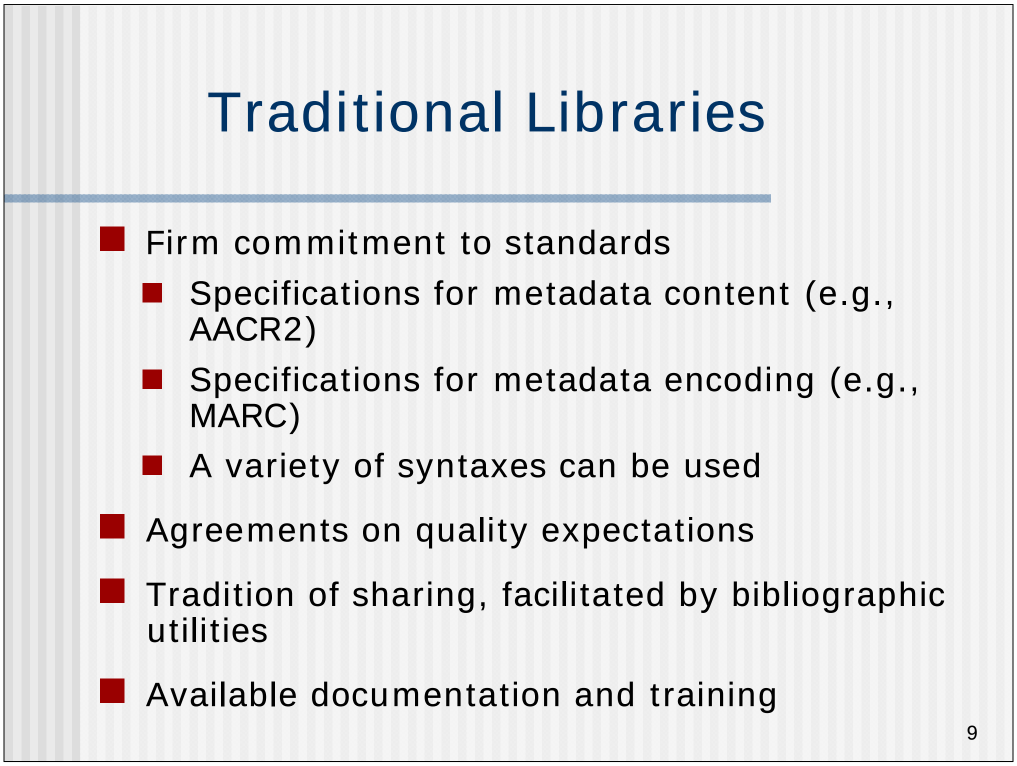
LC is now in the process of developing the World Digital Library.

Google donated $3 million to LC to develop the specifications for a WDL.

This is just the homepage for the WDL prototype. To learn more about WDL prototype the URL is provided.

This is the prototype, the WDL does not yet exist. The point is, we are now very much in a digital environment.

How does the environment affect the creation of metadata?



Note that RDA is developing to replace AACR2 for metadata content; it is significantly different than AACR2, and those differences will be discussed later in the program.

Syntaxes include “classic” MARC and MARCXML

Characteristics of commitment to principles of bibliographic control:

--Sense of quality that is commonly understood

--Notion of completeness for a bibliographic record

--Documentation is ubiquitous– a well-established body of knowledge (we pride ourselves on having a good handle on the documentation)

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

In our current environment, things are evolving.

A Digital Library is to some extent a virtual version of a physical. The “content” is in digital formats and presumed to be available to more than one user at a time.

Although we try our best to “port” our traditional library processing and understandings there are significant differences developing because of the additional challenges of digital content.

Does it revolve around information services, or the web platform?

There is a lot of locally-invented “stuff” involved in many digital libraries, but increasingly standards are being developed for digital libraries.

Some may say that we have so many standards now that they are competing. What may be more interesting is that we have evolving standards that allow the various metadata standards to interoperate.

[Note reference here to OAI-PMH; we’ll talk more about that later]

We have been in dire need of documentation and training for this crossover from catalogers to metadata specialists and that is why this training is taking place.

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

People from libraries are trying to figure out how to operate in the the digital library world.

One thing that creates competition is scale. Compared to published materials, the scale of digital stuff to manage is staggering. As a result, the digital library world is trying to manage with as little human intervention as possible. We want to facilitate people getting at their stuff without increasing the cost. This involves a close examination of what we attempt to do that is based on traditional library practices.

“In the future, the people that are now known as ‘catalogers’ will become ‘ontology midwives.’” – John Unsworth (Dean, GSLIS, University of Illinois, Urbana-Champaign)

The skill set is similar, the management of the information is not.

Players: Unlike traditional libraries, metadata is not necessarily created by catalogers or librarians. In some instances technical specialists take the legacy records and convert them into the standard of their choice.

Goals: Libraries have lost many users to Google. Therefore, libraries and cultural institutions are collaborating with these search engines to sustain our existence and retain users.

Resources: Traditional cataloging is too expensive in this environment where we are overwhelmed with digital information. Funds are low and the demand for resources, be it people and/or money is high. Many understand the value of quality metadata, but in this competitive environment for funds, it depends on where the priorities are placed. We have to change with the times.

Traditionally, we cataloged a book, not the chapter. In this environment granularity issues are broad.

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

We're still trying to find stuff, for example, using identifiers based on the physical world, such as ISBD, ISSN.

There is pressure to do more with less, given the assumption that we will be dealing soon with vast quantities of information.

A lot of early efforts were funded by grants, which did not fund continued support. These days there are few general grants for initial steps in building a digital library, and most still do not support maintenance and ongoing effort.

What *IS* Metadata?

- Some possibilities:
  - Data about data (or data about resources)
  - Structured information that describes, explains, locates, and otherwise makes it easier to retrieve and use an information resource."
  - A management tool
  - Computer-processible, human-interpretable information about digital and non-digital objects

13

"Data about Data" definition largely deprecated (though still seen in documentation).

Second definition from OASIS: a non-profit consortium concerned with open standards

Cataloging is a subset of metadata. Metadata goes beyond the bibliographic record.

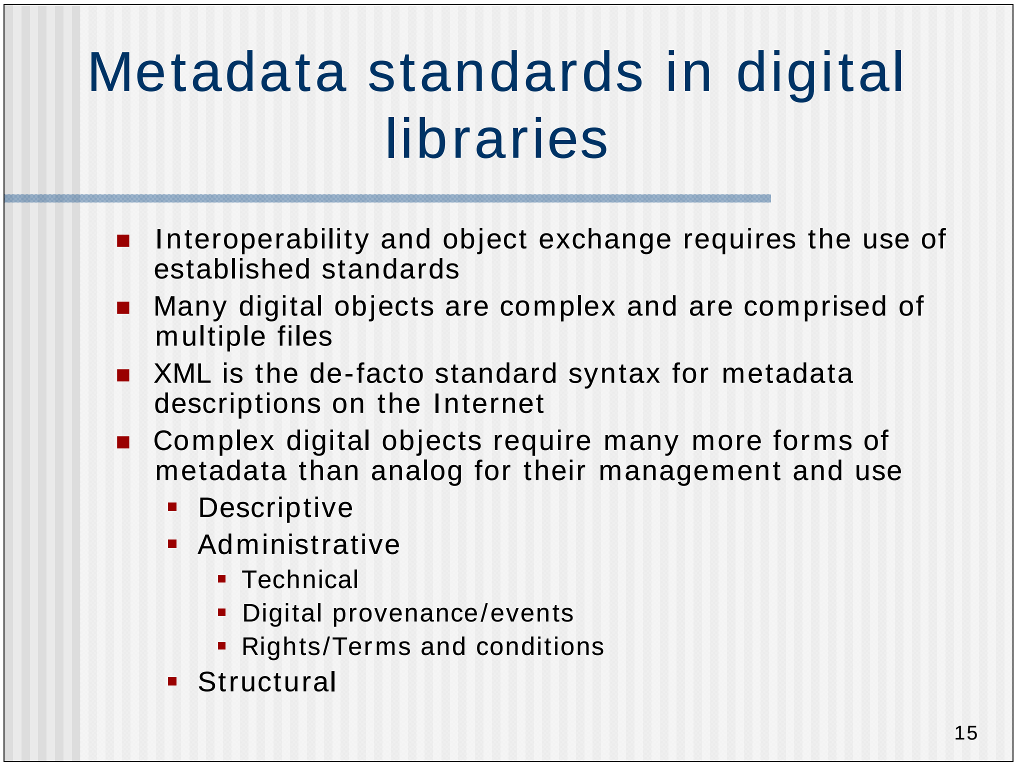
Metadata can be used to manage or manipulate the resources.

Favorite definition of metadata: "**Metadata is cataloging done by men"**

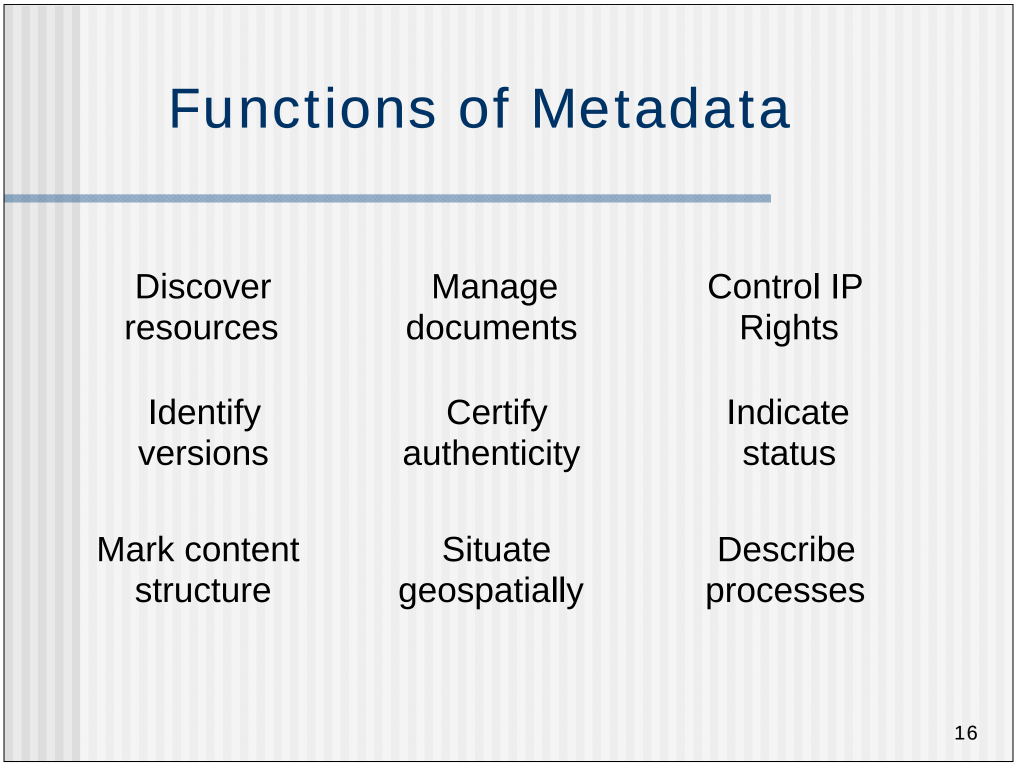
"In moving from dispersed digital collections to interoperable digital libraries, the most important activity we need to focus on is standards… most important is the wide variety of metadata standards [including] descriptive metadata… administrative metadata…, structural metadata, and terms and conditions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

Howard Besser, NYU

14



XML is just the syntax for metadata descriptions. It needs a metadata element set to be used for metadata.



Here are some examples of the functions of metadata.

Different metadata elements are needed to perform different tasks.

Some examples: Author, title and subject support the function of discovery;

Copyright information supports the control of intellectual property rights; geospatial coordinates support situating geospatially.

Types of metadata

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

17

Metadata can take various forms. This categorization is consistent with the National Information Standard Organization's (NISO) publication "Understanding Metadata".

Descriptive: the metadata that is well established that we all know and love. It supports user tasks such as resource discovery, identification, acquisition, etc. Examples are author, subject.

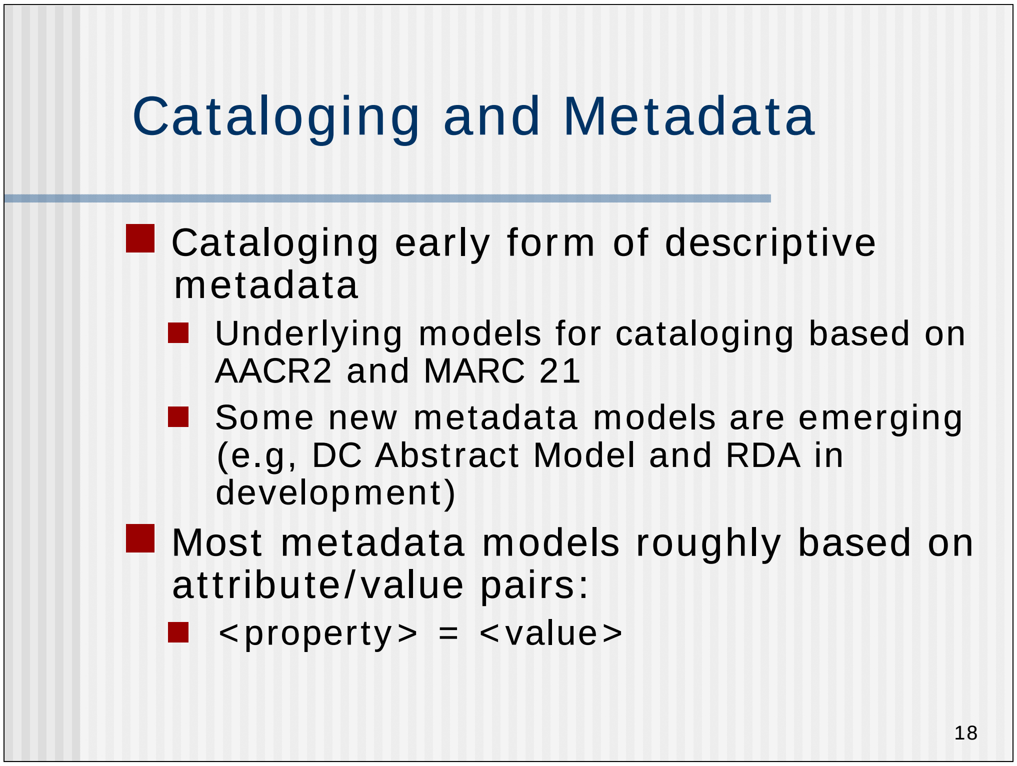
Administrative metadata: Metadata to manage the object. It has subcategories of technical (technical characteristics about the object), digital provenance (actions that have been performed on the object and source and custody of the object), rights (information about access and use of the object)

Preservation metadata supports long-term preservation; may fall into one of the other 3 categories.

Structural: Information about an object's physical structure; it ties the components of a complex or compound resource together and makes the whole usable (example is paging for digital books when each page is an image).

Meta-metadata is information about the metadata itself; who created it, when, where it came from, when it was updated.

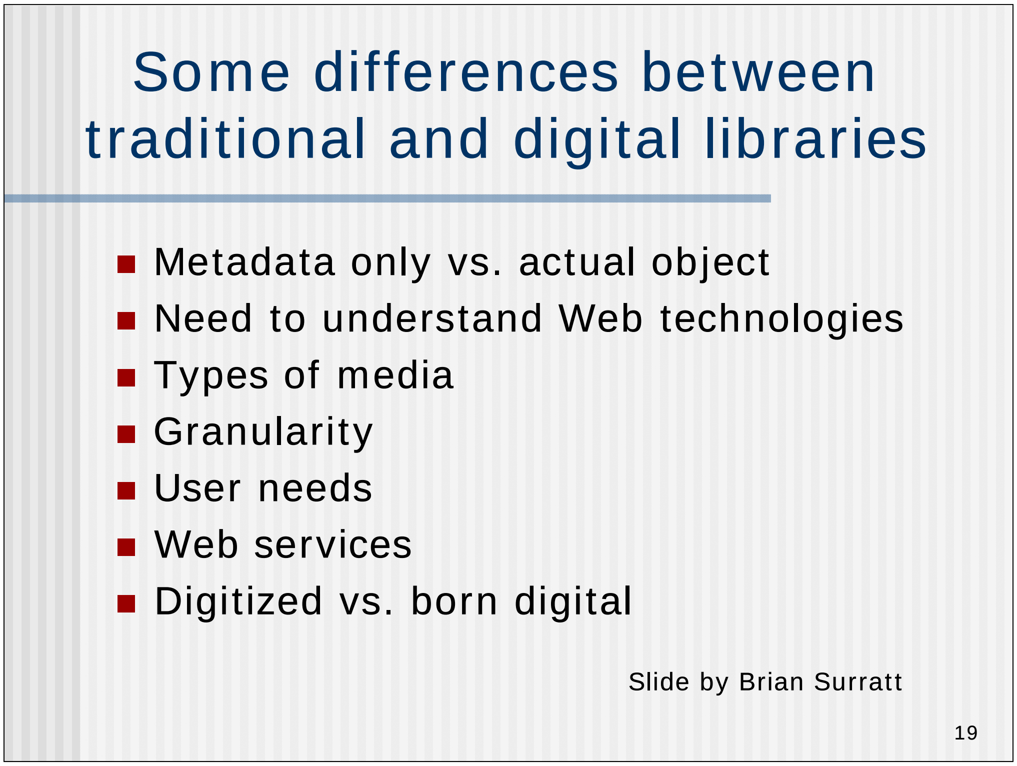
Types of metadata will be covered in detail later in sessions 2, 3, and 4.



Although some people make a distinction between "cataloging" and "metadata", cataloging is really just another term for metadata with a particular focus. Its origins were for a library "catalog", but the content of a catalog record really consists of multiple metadata fields for resource description.

The movement to create FRBR (Functional Requirements of Bibliographic Records) relationships helped to break down cataloging practice into fundamental principles.

Dublin Core Abstract Model: to get at the underlying relationships that drive metadata description.



•Traditional library catalogs are only designed to hold the metadata, not the digital item. When you are working with digital libraries, you are working with the actual object (full-text, audio, video, data) and you have to think about the relationship of the two.

•When working with metadata one has to have a much greater understanding of basic web technologies: servers, TCP-IP, DNS, HTML, etc.

•The types of media that can be delivered over the web is exploding. Many OPACs primarily describe textual objects, but digital libraries contain all types of media: video, audio, images, text, and these can be integrated together

•Granularity – the internet is a good medium for storing and delivering smaller chunks of information: individual photos, articles, small audio and video files. The OPAC is built to describe monographs, journal titles, and collections. There is a lot more opportunity/pressure to describe things at a lower level of granularity in digital libraries.

•User needs – web design is influenced by user behavior. In libraries it has been the other way around, we have designed a complex, controlled system and then we try to teach people to use it.

•Web services. Digital libraries are not static things, they are like books with wings. Web services are the things you can do with the metadata.

•Born digital vs. digitized. The techniques for managing born-digital vs. retrospectively digitized collections are different. They are both challenging in different ways.

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

**Thanks to Marty Kurth for these insights*

20

It is necessary to think about things at multiple levels. There is an analogy with the archival process.

Questions to ask when selecting metadata standards

- 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 will they 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?

21

Think about this slide as we go through the course. These are questions to consider when selecting appropriate metadata standards.

Some things to keep in mind:

Type of material:

•Format (photographs, documents, oral histories, raw data)

•Complexity of relationships (is it archival in nature and do individual items make sense out of a broader context; is there a hierarchical relationship like photo albums)

•Are there multiple versions that need to be displayed

•Who is the intended audience and their expectations and usage

•The availability of tools can be an issue and may affect the choice of standard. For instance it would be difficult to use METS as a container without the ability to generate a METS object from a tool, since it is not intended to be done by hand. Fortunately there are a number of open source tools available.

See instructor's notes for details on this exercise.

Invite the audience to comment on how these sites are structured.

When you look at a digital site, look at how they categorize the materials. For example: "Search" under keyword, title, creator; "Media" is broken down into article, book, CD, chapter, film, website. Look at these sites not only as a user but also as the developer.


ALCTS
2. Specific metadata standards: descriptive
Metadata Standards and Applications Workshop

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 and the community that developed and use them
- Learn about some relationship models used in descriptive metadata standards

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...)
- Relationship models

3

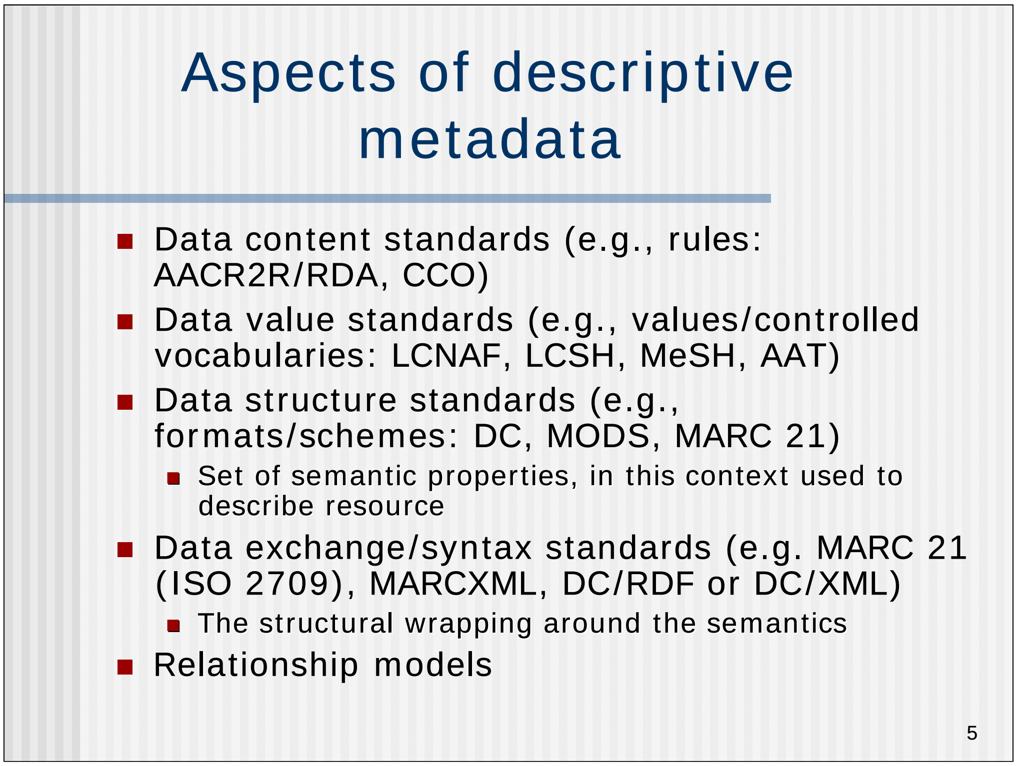
Descriptive metadata

- 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

Of course descriptive metadata has been around for a very long time, as we have been describing analog objects for centuries. The term descriptive metadata was coined about 20 years ago, but essentially refers to bibliographic descriptions. As more and more resources have become digital, its importance outside the library community has increased, and other communities have discovered its utility.

It is the most standardized…



It is important to understand the scope of different types of descriptive metadata and how the pieces fit together in creating descriptions of resources.

A distinction is made here between the rules for populating elements (data content standards) from controlled vocabularies used as values in specific fields in metadata (data value standards). What we are referring to as data structure standards may just be element sets with definitions, or may have more detailed instructions that indicate additional structure. Relationship models are included here, since they inform the development of many of the other categories of descriptive metadata. Syntax standards are used for data exchange and should be considered as independent of the data structure standard, although some are associated together in practice.

Much of these categories also apply to technical metadata (which we will discuss in session 3), although descriptive is more well developed.

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

RDA work is undergoing development, and its focus is still being debated to some extent. Best practices and guidelines may be formal or informal documents that are not standardized in the same way as other standards, but they may accompany standards and provide useful information about implementation. They supplement formal content rules; examples are LCRIs and local policies or guidelines

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

Controlled vocabularies may be machine readable or not, although for maintenance purposes this would be desirable.

There are many other examples of controlled vocabularies developed by different groups and communities that may be used in descriptive metadata.

We will cover this area in more detail in session 5 on applying metadata standards.

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

People also refer to data structure standards as “schemas” or “schemes”. This is not the same as “XML schema”, which is tied to a particular syntax and is the structural “template” that allows XML data to be validated.

Syntax allows metadata to be shared, move around, and so forth

Schemas are essential for understanding “what it is” that the metadata is supposed to be doing.

Data Structure Standards: Examples

- MARC 21 (http://www.loc.gov/marc/)
- Dublin Core (http://dublincore.org)
- MODS (www.loc.gov/standards/mods/)
- IEEE-LOM (http://ltsc.ieee.org/wg12/)
- ONIX (http://www.editeur.org/onix.html)
- EAD (http://www.loc.gov/ead/)

9

Dublin Core was developed to apply to a wide variety of digital objects.

MODS= Metadata Object Description Schema; was developed by the Library of Congress as an alternative to MARC for the description of online resources.

IEEE-LOM (eye-triple-ee lahm) is a schema for educational materials.

ONIX=Online Information Exchange; ONIX for Books is what is used by Amazon, BN and other online bookstores. There is an ONIX for Serials as well.

EAD=Encoded Archival Description, is used to describe archival collections.

Data Structure Standards: Examples, cont.

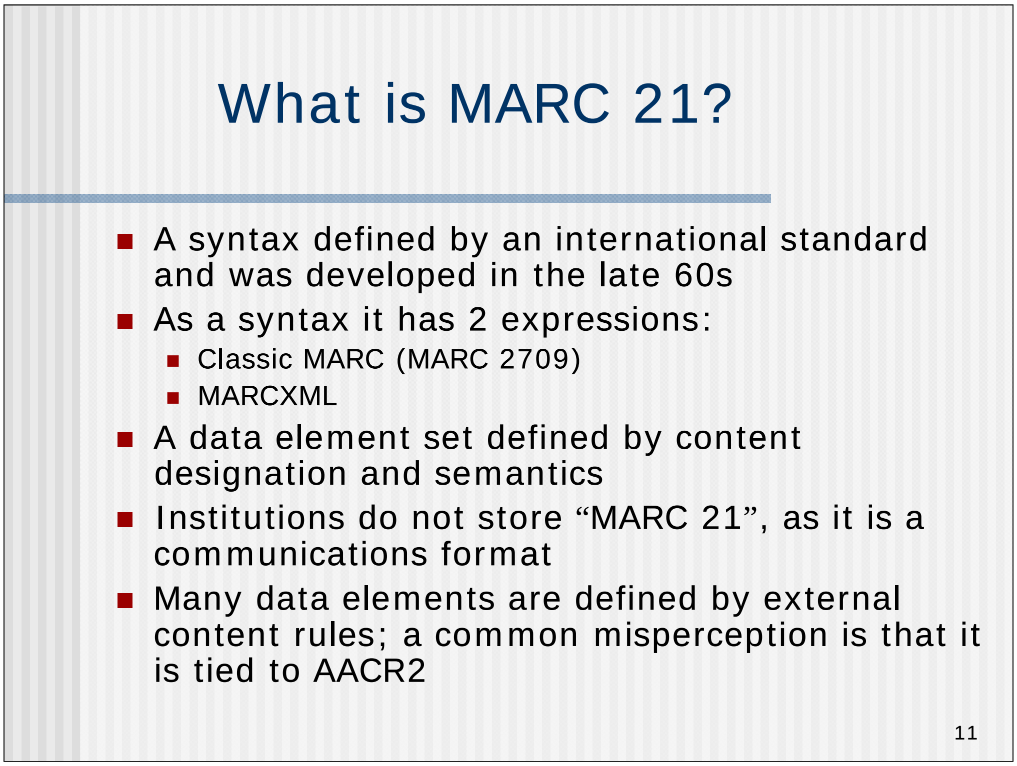
- VRA Core (http://www.vraweb.org/projects/vracore4/)
- PBCore (http://www.pbcore.org/)
- TEI (http://www.tei-c.org/index.xml)

10

CDWA/VRA Core=Categories for Description of Works of Art/Visual Resources Association Core elements; developed to describe digitized works of art

PBCore=Public Broadcasting Core; an element set built on Dublin Core, developed by the Corporation for Public Broadcasting to describe moving image objects

TEI (Text Encoding Initiative) has been used since 1994 to encode texts, including parts of speech, important quotes, names, numbers and dates. The TEI Header typically includes standard descriptive metadata such as title, edition, extent, and so on.



Many data elements are defined by external content rules; a common misperception is that it is tied to AACR2.

02158cam 2200349Ia
4500001001300000003000600013005001700019006001900036006001900055007001500074008004
1000890400020001300200015001500430021001650490009001862450119001952460025
0031426000650033953800300040450600380043453601530047252007640062550500940138950000
86014836000049015696500040016186510039016586510023016977000026017208560050017469
94001201796 ocm56835268 OCoLC 20060118051017.0 m d szx w s 0 2 cr mn---
------ 041028m20049999vau st 000 0 eng d aVA@ cVA@ dOCLCQ a0813922917
an-us--- an-us-va aVA@@ 04 aThe Dolley Madison digital edition h[electronic resource]
: bletters 1788-June 1836 / cedited by Holly C. Shulman. 1 iAlso known as: aDMDE
aCharlottesville, Va. : bUniversity of Virginia Press, c2004- aMode of access: Internet.
aSubscription required for access. aRotunda editions are made possible by generous grants
from the Andrew W. Mellon Foundation and the President's Office of the University of Virginia.
aDolley Payne Madison was the most important First Lady of the nineteenth century. The DMDE will
be the first-ever complete edition of all of her known correspondence, gathered in an XML-based
archive. It will ultimately include close to 2,500 letters. From the scattered correspondence were
gathered letters that have never been previously published. The range and scope of the collection
makes this edition an important scholarly contribution to the literature of the early republic, women's
history, and the institution of the First Lady. These letters present Dolley Madison's trials and triumphs
and make it possible to gain admittance to her mind and her private emotions and to understand the
importance of her role as the national capital's First Lady. 0 aGeneral introduction -- Biographical
introduction -- Introduction to the digital edition. aTitle from the opening screen; description based
on the display of Oct. 21, 2004. 10 aMadison, Dolley, d1768-1849 vCorrespondence.
0 aPresidents' spouses zUnited States. 0 aUnited States xHistory y1801-1809.
0 aVirginia xHistory. 1 aShulman, Holly
Cowan. 40 uhttp://rotunda.upress.virginia.edu:8100/dmde/ aC0 bVA@

A raw MARC record -- this is how MARC systems exchange data. It shows both data elements and MARC syntax.

Animation (on click):

#1 -- Title will appear (Dolley Madision …)

How does the computer find the place in the record for the title?

#2 -- identify the 245 tag

#3 -- 119 -- that number of characters that tag contains (in this example, the title is 119 characters long)

#4 -- 195 -- from the start point, count 195 characters to reach the beginning of the 245 tag.

ask: you can tell from looking, that if the title is 119 characters long, it can't be placed 195 characters from the beginning. There are too many characters between 02158 … and the beginning of the title. how do you know where to start counting

#5 -- 00349 -- start counting at position 349.

#6 -- 5 (this is position 349. A computer would start counting here). Count 195 characters. Reach the beginning of the 245. Count 119 characters & reach the end of the title tag.

start over & show 260 example

clicks #7-10, will make the previous example disappear.

#11 -- identify 260 tag

#12 – 339 -- count 339 characters from the start position (step #6 above)

#13 -- 0065 -- go 65 characters & you will find:

#14 -- the start of the 260

MARC 21 Scope

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

13

MARC formats for Classification data and Community Information are not as widely used as the others.

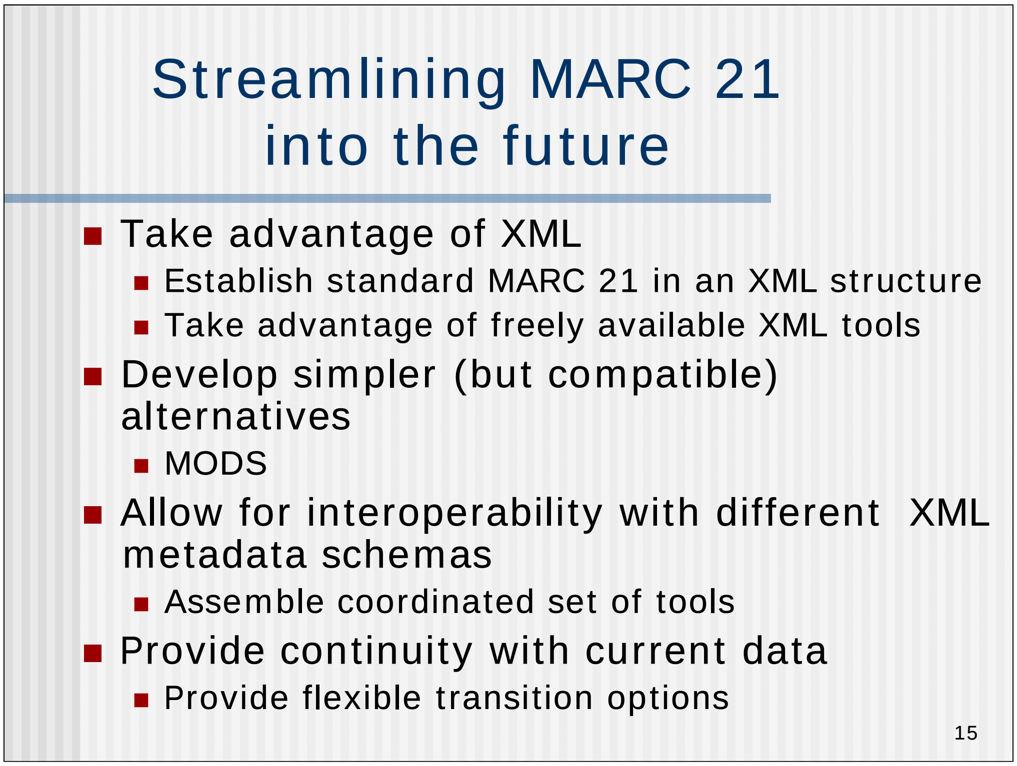
MARC 21 implementation

- 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

14

A brief review of MARC is important to understand where we're coming from and where we might be going in the future.

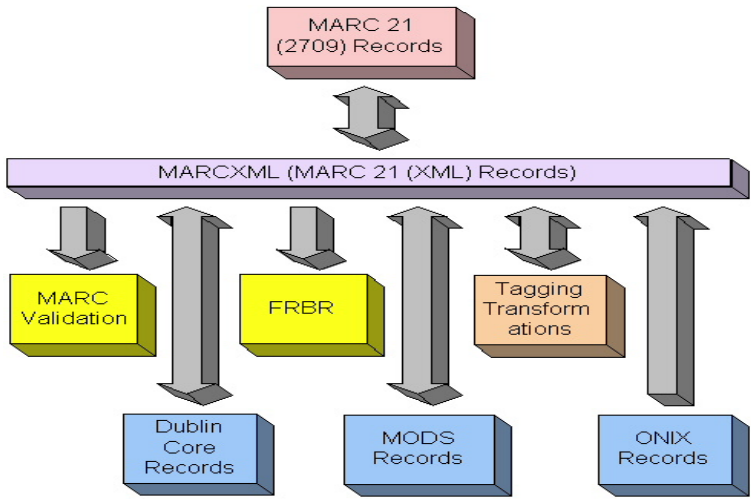
Although there is some talk about MARC not surviving into the future, a number of countries have recently given up their national formats and replaced them with standard MARC 21.



Although some have said that MARC must die and that it is not appropriate in today's environment, it is important to keep in mind the large investment in systems and records based on MARC. LC is exploring streamlining MARC and evolving it into the future so that institutions don't have to throw out all their data and start with something new.

We are introducing the use of XML with descriptive metadata. Details about the XML syntax itself will be given in session 4.

MARC 21 evolution to XML



MARCXML is a syntactical conversion from classic MARC; once in XML other tools can be applied for further transformations into different metadata schemes/formats. Here are some examples of transformations.

Note that ONIX is only going one way, but it is expected that most of its use in relation to MARC will be the exchange of an initial record that could be a basis for a library catalog record.

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

17

The first link shows a record with MARC tags (from LC's catalog).

The second shows how the syntax using the MARCXML schema. Note that there are a minimum of elements and the tag itself is contained as a value of an attribute. Thus, the schema does not need to be changed when new data elements are added to MARC and can be used for any of the five MARC 21 formats.

We will look at MARCXML again in session 4 on syntax.

Printouts are included in the manual.

The Library of Congress

>> Go to Library of Congress Authorities

LIBRARY OF CONGRESS ONLINE CATALOG

Help | New Search | Search History | Headings List | Titles List | Request an Item | Account Status | Start Over

DATABASE: Library of Congress Online Catalog
YOU SEARCHED: Command = 010a 85753651
SEARCH RESULTS: Displaying 1 of 1.

Previous Next

Brief Record | Subjects/Content | Full Record | MARC Tags

3 Viennese arias : for soprano, obbligato clarinet in B flat, and piano /...

LC Control No.: 85753651
LCCN Permalink: http://lccn.loc.gov/85753651

000 01917ccm a2200409 a 450
001 5594130
005 19950601141653.9
008 850813s1984 enkopa z n ita
035 __ |**9** (DLC) 85753651
906 __ |**a** 7 |**b** cbc |**c** orignew |**d** 3 |**e** ncip |**f** 19 |**g** y-genmusic
010 __ |**a** 85753651
028 32 |**a** N.M. 275 |**b** Nova Music
040 __ |**a** DLC |**c** DLC |**d** DLC
041 1_ |**a** ita |**e** itaeng |**h** ita
048 __ |**b** va01 |**a** wc01 |**a** ka01
050 00 |**a** M1506 |**b** .A14 1984
245 00 |**a** 3 Viennese arias : |**b** for soprano, obbligato clarinet in B flat, and piano / |**c** G.B. Bononcini and Emperor Joseph I ; edited by Colin Lawson.
260 __ |**a** London : |**b** Nova Music, |**c** c1984.
300 __ |**a** 1 score (12 p.) + 2 parts ; |**c** 31 cm.
440 _0 |**a** Music for voice and instrument
500 __ |**a** Opera excerpts.
500 __ |**a** Acc. arr. for piano; obbligato for the 2nd-3rd excerpts originally for chalumeau.
500 __ |**a** Italian words.
500 __ |**a** Cover title.
500 __ |**a** The 1st excerpt composed for inclusion in M.A. Ziani's Chilonida.
500 __ |**a** Texts with English translations on cover p. [2].
505 0_ |**a** Tutto in pianto il cor struggete / Emperor Joseph I -- E sempre inquieto quel

core infelice : from Endimione / G. Bononcini -- L'adorata genitrice : from Muzio [i.e. Mutio] Scevola / G. Bononcini.

650 _0 |**a** Operas |**x** Excerpts, Arranged |**x** Scores and parts.

650 _0 |**a** Songs (High voice) with instrumental ensemble |**x** Scores and parts.

700 1_ |**a** Lawson, Colin |**q** (Colin James)

700 02 |**a** Joseph |**b** I, |**c** Holy Roman Emperor, |**d** 1678-1711. |**t** Tutto in pianto il cor struggete; |**o** arr. |**f** 1984.

700 12 |**a** Bononcini, Giovanni, |**d** 1670-1747. |**t** Endimione. |**p** E sempre inquieto quel core infelice; |**o** arr. |**f** 1984.

700 12 |**a** Bononcini, Giovanni, |**d** 1670-1747. |**t** Mutio Scevola. |**p** Adorata genitrice; |**o** arr. |**f** 1984.

740 0_ |**a** Three Viennese arias.

740 0_ |**a** Viennese arias.

953 __ |**a** TA28

991 __ |**b** c-Music |**h** M1506 |**i** .A14 1984 |**t** Copy 1 |**w** MUSIC

CALL NUMBER: M1506 .A14 1984
Copy 1
-- Request in: Performing Arts Reading Room (Madison, LM113)
-- Status: Not Charged

Previous Next

Save, Print or Email Records (View Help)		
Select Download Format:	Text (Brief Information)	Press to SAVE or PRINT
Email Text (Full Info) to:		Press to SEND EMAIL

Help - Search - History - *Headings* - Titles - Request - Account - Exit

---

The Library of Congress
URL: http://www.loc.gov/
Mailing Address:
101 Independence Ave, S.E.
Washington, DC 20540

Catalog/authority record errors?
Use our Error Report Form
Questions about searching?
Ask a Librarian

Library of Congress Online Catalog
URL: http://catalog.loc.gov/
Library of Congress Authorities
URL: http://authorities.loc.gov/

```
  <?xml version="1.0" ?>
- <zs:searchRetrieveResponse xmlns:zs="http://www.loc.gov/zing/srw/">
    <zs:version>1.1</zs:version>
    <zs:numberOfRecords>1</zs:numberOfRecords>
  - <zs:records>
    - <zs:record>
        <zs:recordSchema>info:srw/schema/1/marcxml-v1.1</zs:recordSchema>
        <zs:recordPacking>xml</zs:recordPacking>
      - <zs:recordData>
        - <record xmlns="http://www.loc.gov/MARC21/slim">
            <leader>01917ccm 2200409 a 4500</leader>
            <controlfield tag="001">5594130</controlfield>
            <controlfield tag="005">19950601141653.9</controlfield>
            <controlfield tag="008">850813s1984 enkopa z n ita</controlfield>
          - <datafield tag="035" ind1="" ind2="">
              <subfield code="9">(DLC) 85753651</subfield>
            </datafield>
          - <datafield tag="906" ind1="" ind2="">
              <subfield code="a">7</subfield>
              <subfield code="b">cbc</subfield>
              <subfield code="c">orignew</subfield>
              <subfield code="d">3</subfield>
              <subfield code="e">ncip</subfield>
              <subfield code="f">19</subfield>
              <subfield code="g">y-genmusic</subfield>
            </datafield>
          - <datafield tag="010" ind1="" ind2="">
              <subfield code="a">85753651</subfield>
            </datafield>
          - <datafield tag="028" ind1="3" ind2="2">
              <subfield code="a">N.M. 275</subfield>
              <subfield code="b">Nova Music</subfield>
            </datafield>
          - <datafield tag="040" ind1="" ind2="">
              <subfield code="a">DLC</subfield>
              <subfield code="c">DLC</subfield>
              <subfield code="d">DLC</subfield>
            </datafield>
          - <datafield tag="041" ind1="1" ind2="">
              <subfield code="a">ita</subfield>
              <subfield code="e">itaeng</subfield>
              <subfield code="h">ita</subfield>
            </datafield>
          - <datafield tag="048" ind1="" ind2="">
              <subfield code="b">va01</subfield>
              <subfield code="a">wc01</subfield>
              <subfield code="a">ka01</subfield>
            </datafield>
          - <datafield tag="050" ind1="0" ind2="0">
              <subfield code="a">M1506</subfield>
              <subfield code="b">.A14 1984</subfield>
            </datafield>
```

```
- <datafield tag="245" ind1="0" ind2="0">
    <subfield code="a">3 Viennese arias :</subfield>
    <subfield code="b">for soprano, obbligato clarinet in B flat,
      and piano /</subfield>
    <subfield code="c">G.B. Bononcini and Emperor Joseph I ;
      edited by Colin Lawson.</subfield>
  </datafield>
- <datafield tag="260" ind1="" ind2="">
    <subfield code="a">London :</subfield>
    <subfield code="b">Nova Music,</subfield>
    <subfield code="c">c1984.</subfield>
  </datafield>
- <datafield tag="300" ind1="" ind2="">
    <subfield code="a">1 score (12 p.) + 2 parts ;</subfield>
    <subfield code="c">31 cm.</subfield>
  </datafield>
- <datafield tag="440" ind1="" ind2="0">
    <subfield code="a">Music for voice and instrument</subfield>
  </datafield>
- <datafield tag="500" ind1="" ind2="">
    <subfield code="a">Opera excerpts.</subfield>
  </datafield>
- <datafield tag="500" ind1="" ind2="">
    <subfield code="a">Acc. arr. for piano; obbligato for the 2nd-
      3rd excerpts originally for chalumeau.</subfield>
  </datafield>
- <datafield tag="500" ind1="" ind2="">
    <subfield code="a">Italian words.</subfield>
  </datafield>
- <datafield tag="500" ind1="" ind2="">
    <subfield code="a">Cover title.</subfield>
  </datafield>
- <datafield tag="500" ind1="" ind2="">
    <subfield code="a">The 1st excerpt composed for inclusion in
      M.A. Ziani's Chilonida.</subfield>
  </datafield>
- <datafield tag="500" ind1="" ind2="">
    <subfield code="a">Texts with English translations on cover p.
      [2].</subfield>
  </datafield>
- <datafield tag="505" ind1="0" ind2="">
    <subfield code="a">Tutto in pianto il cor struggete / Emperor
      Joseph I -- E sempre inquieto quel core infelice : from
      Endimione / G. Bononcini -- L'adorata genitrice : from
      Muzio [i.e. Mutio] Scevola / G. Bononcini.</subfield>
  </datafield>
- <datafield tag="650" ind1="" ind2="0">
    <subfield code="a">Operas</subfield>
    <subfield code="x">Excerpts, Arranged</subfield>
    <subfield code="x">Scores and parts.</subfield>
  </datafield>
- <datafield tag="650" ind1="" ind2="0">
```

```
      <subfield code="a">Songs (High voice) with instrumental
        ensemble</subfield>
      <subfield code="x">Scores and parts.</subfield>
    </datafield>
  - <datafield tag="700" ind1="1" ind2="">
      <subfield code="a">Lawson, Colin</subfield>
      <subfield code="q">(Colin James)</subfield>
    </datafield>
  - <datafield tag="700" ind1="0" ind2="2">
      <subfield code="a">Joseph</subfield>
      <subfield code="b">I,</subfield>
      <subfield code="c">Holy Roman Emperor,</subfield>
      <subfield code="d">1678-1711.</subfield>
      <subfield code="t">Tutto in pianto il cor struggete;</subfield>
      <subfield code="o">arr.</subfield>
      <subfield code="f">1984.</subfield>
    </datafield>
  - <datafield tag="700" ind1="1" ind2="2">
      <subfield code="a">Bononcini, Giovanni,</subfield>
      <subfield code="d">1670-1747.</subfield>
      <subfield code="t">Endimione.</subfield>
      <subfield code="p">E sempre inquieto quel core
        infelice;</subfield>
      <subfield code="o">arr.</subfield>
      <subfield code="f">1984.</subfield>
    </datafield>
  - <datafield tag="700" ind1="1" ind2="2">
      <subfield code="a">Bononcini, Giovanni,</subfield>
      <subfield code="d">1670-1747.</subfield>
      <subfield code="t">Mutio Scevola.</subfield>
      <subfield code="p">Adorata genitrice;</subfield>
      <subfield code="o">arr.</subfield>
      <subfield code="f">1984.</subfield>
    </datafield>
  - <datafield tag="740" ind1="0" ind2="">
      <subfield code="a">Three Viennese arias.</subfield>
    </datafield>
  - <datafield tag="740" ind1="0" ind2="">
      <subfield code="a">Viennese arias.</subfield>
    </datafield>
  - <datafield tag="953" ind1="" ind2="">
      <subfield code="a">TA28</subfield>
    </datafield>
  - <datafield tag="991" ind1="" ind2="">
      <subfield code="b">c-Music</subfield>
      <subfield code="h">M1506</subfield>
      <subfield code="i">.A14 1984</subfield>
      <subfield code="t">Copy 1</subfield>
      <subfield code="w">MUSIC</subfield>
    </datafield>
  </record>
</zs:recordData>
```

```
      <zs:recordPosition>1</zs:recordPosition>
    </zs:record>
  </zs:records>
</zs:searchRetrieveResponse>
```

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

19

Use of XML schema allows for flexibility and availability of freely available tools

MODS is a simpler view of MARC using language tags rather than numeric tags, so is somewhat more user friendly. As with other metadata schemes, the tag names could be easily translated into other languages using common XML tools. In cases where there are redundancies in MARC (e.g. genre values in 008 in additional to in variable fields), the data has been brought together into one data element with an “authority” attribute used to specify the controlled vocabulary.

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

20

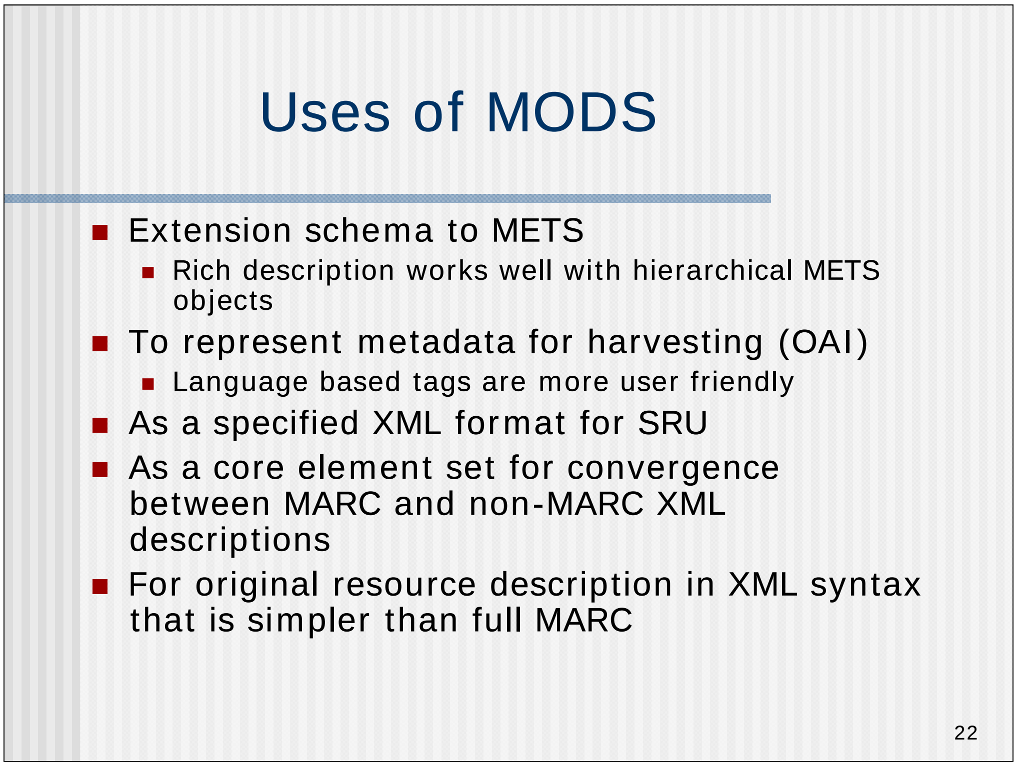
All MODS elements allow for specifying a language, script, or transliteration at the element level.

In MARC you can only specify language at the record level. XML by definition uses Unicode.

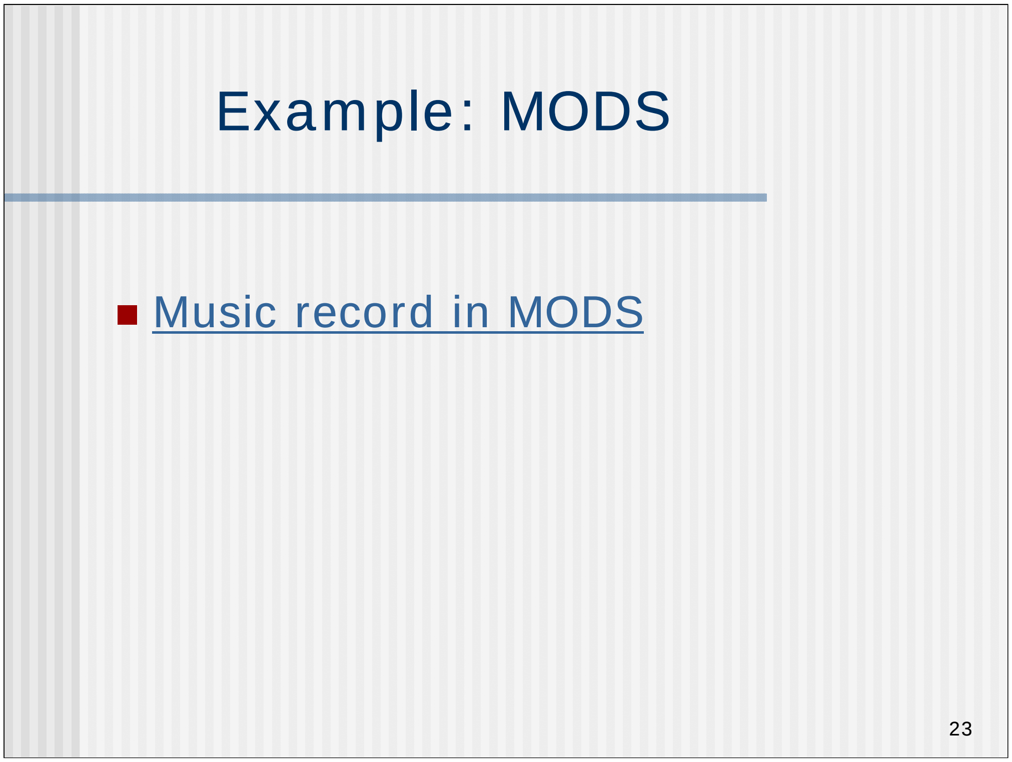
Advantages of MODS

- 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

21



Originally MODS was developed because of the need for a richer description than Dublin Core but more user friendly and simple than full MARC 21, especially for the emerging initiatives that required XML output (METS, OAI, SRU)



Click on the link to see the full record in MODS. The example is also available on the following page in the manual.

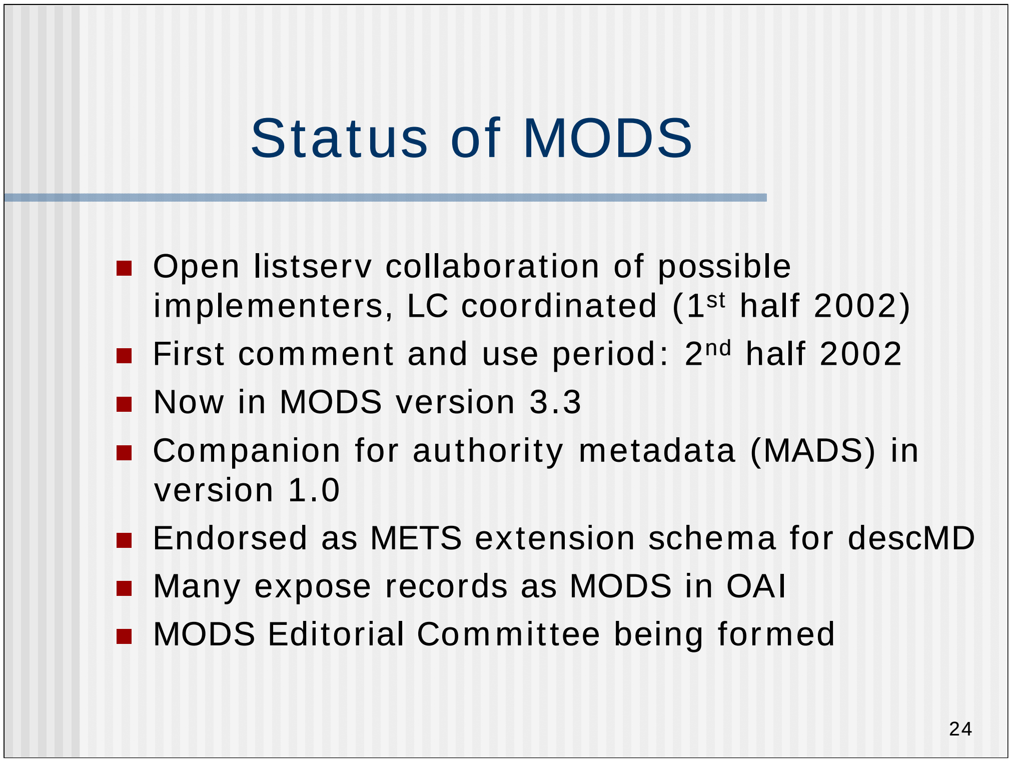
This is the same record that we saw in MARC and MARCXML.

Note the use of authority attributes to specify controlled vocabularies. Note also the use of relatedItem with a type="constituent". This allows for more detailed descriptions of constituent parts of the item, which are listed in the table of contents. We will see this flexibility to describe at any level of granularity again when we look at its application in digital libraries.

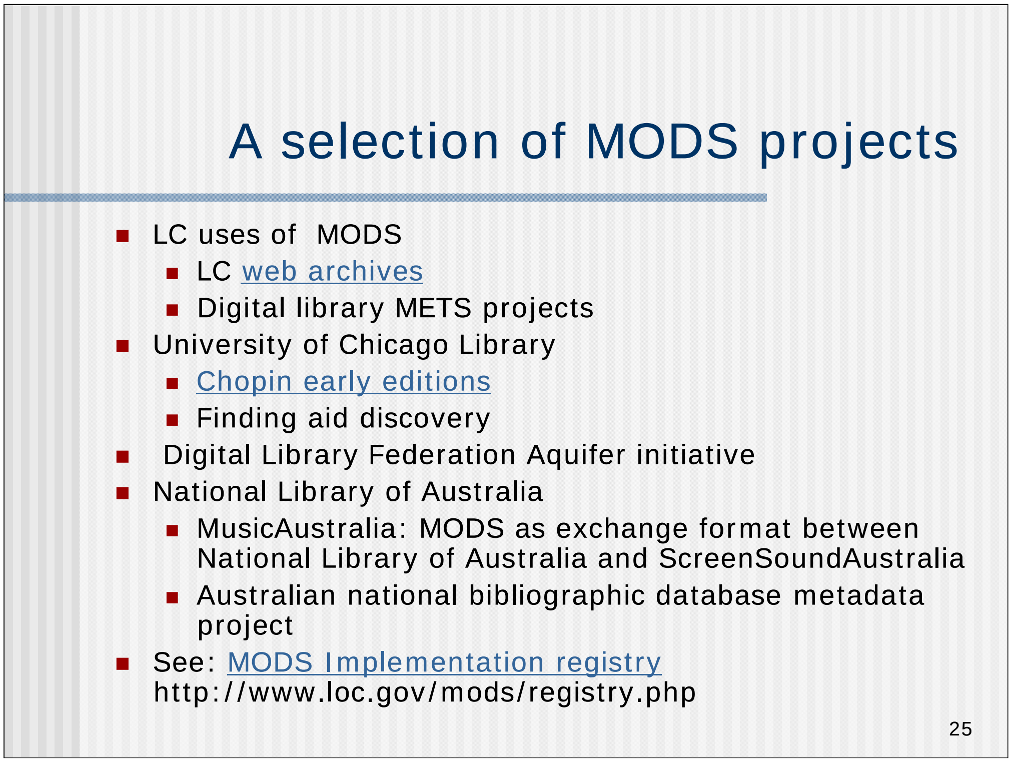
```
<?xml version="1.0" ?>
- <zs:searchRetrieveResponse xmlns:zs="http://www.loc.gov/zing/srw/">
    <zs:version>1.1</zs:version>
    <zs:numberOfRecords>1</zs:numberOfRecords>
  - <zs:records>
    - <zs:record>
        <zs:recordSchema>info:srw/schema/1/mods-v3.0</zs:recordSchema>
        <zs:recordPacking>xml</zs:recordPacking>
      - <zs:recordData>
        - <mods xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xmlns="http://www.loc.gov/mods/v3" version="3.0"
            xsi:schemaLocation="http://www.loc.gov/mods/v3
            http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
          - <titleInfo>
              <title>3 Viennese arias</title>
              <subTitle>for soprano, obbligato clarinet in B flat, and
                piano</subTitle>
            </titleInfo>
          - <name type="personal">
              <namePart>Lawson, Colin (Colin James)</namePart>
            </name>
            <typeOfResource>notated music</typeOfResource>
          - <originInfo>
            - <place>
                <placeTerm type="code"
                  authority="marccountry">enk</placeTerm>
              </place>
            - <place>
                <placeTerm type="text">London</placeTerm>
              </place>
              <publisher>Nova Music</publisher>
              <dateIssued>c1984</dateIssued>
              <dateIssued encoding="marc">1984</dateIssued>
              <issuance>monographic</issuance>
            </originInfo>
          - <language>
              <languageTerm authority="iso639-2b"
                type="code">ita</languageTerm>
            </language>
          - <language>
              <languageTerm authority="iso639-2b"
                type="code">eng</languageTerm>
            </language>
          - <physicalDescription>
              <form authority="marcform">print</form>
              <extent>1 score (12 p.) + 2 parts ; 31 cm.</extent>
            </physicalDescription>
            <tableOfContents>Tutto in pianto il cor struggete / Emperor
              Joseph I -- E sempre inquieto quel core infelice : from
              Endimione / G. Bononcini -- L'adorata genitrice : from Muzio
              [i.e. Mutio] Scevola / G. Bononcini.</tableOfContents>
            <note type="statement of responsibility">G.B. Bononcini and
              Emperor Joseph I ; edited by Colin Lawson.</note>
```

```
<note>Opera excerpts.</note>
<note>Acc. arr. for piano; obbligato for the 2nd-3rd excerpts
  originally for chalumeau.</note>
<note>Italian words.</note>
<note>Cover title.</note>
<note>The 1st excerpt composed for inclusion in M.A. Ziani's
  Chilonida.</note>
<note>Texts with English translations on cover p. [2].</note>
- <subject authority="lcsh">
    <topic>Operas</topic>
    <topic>Excerpts, Arranged</topic>
    <topic>Scores and parts</topic>
  </subject>
- <subject authority="lcsh">
    <topic>Songs (High voice) with instrumental
      ensemble</topic>
    <topic>Scores and parts</topic>
  </subject>
  <classification authority="lcc">M1506 .A14 1984</classification>
- <relatedItem type="series">
  - <titleInfo>
      <title>Music for voice and instrument</title>
    </titleInfo>
  </relatedItem>
- <relatedItem type="constituent">
  - <titleInfo>
      <title>Tutto in pianto il cor struggete; arr. 1984</title>
    </titleInfo>
  - <name type="personal">
      <namePart>Joseph</namePart>
      <namePart type="termsOfAddress">I, Holy Roman
        Emperor</namePart>
      <namePart type="date">1678-1711</namePart>
    </name>
  </relatedItem>
- <relatedItem type="constituent">
  - <titleInfo>
      <title>Endimione. arr. 1984</title>
      <partName>E sempre inquieto quel core infelice; arr.
        1984</partName>
    </titleInfo>
  - <name type="personal">
      <namePart>Bononcini, Giovanni,</namePart>
      <namePart type="date">1670-1747</namePart>
    </name>
  </relatedItem>
- <relatedItem type="constituent">
  - <titleInfo>
      <title>Mutio Scevola. arr. 1984</title>
      <partName>Adorata genitrice; arr. 1984</partName>
    </titleInfo>
  - <name type="personal">
```

```
            <namePart>Bononcini, Giovanni,</namePart>
            <namePart type="date">1670-1747</namePart>
          </name>
        </relatedItem>
        <identifier type="lccn">85753651</identifier>
        <identifier type="music publisher">N.M. 275 Nova
          Music</identifier>
      - <recordInfo>
          <recordContentSource
            authority="marcorg">DLC</recordContentSource>
          <recordCreationDate
            encoding="marc">850813</recordCreationDate>
          <recordChangeDate
            encoding="iso8601">19950601141653.9</recordChangeDate>
          <recordIdentifier>5594130</recordIdentifier>
        </recordInfo>
      </mods>
    </zs:recordData>
    <zs:recordPosition>1</zs:recordPosition>
  </zs:record>
 </zs:records>
</zs:searchRetrieveResponse>
```



MODS seems to have become widely used for digital library projects, especially when there is need to interact with other library records, particularly MARC. Using MODS, the richness of a MARC description is not sacrificed to a simple set of data elements. Some of the granular tagging may be lost in conversion to MODS from MARC, but not much data is.



LC is using a limited set of MODS elements to describe archived websites (click on LC web archives), including a search and browse interface.

The University of Chicago uses METS with MODS for various projects. We will look at Chopin early editions in session 7.

The DLF Aquifer Initiative uses MODS, which promotes distributed digital library content in the area of American social history.

The National Library of Australia uses MODS as a common format between the national bibliographic database and non-MARC metadata in ScreenSound Australia.

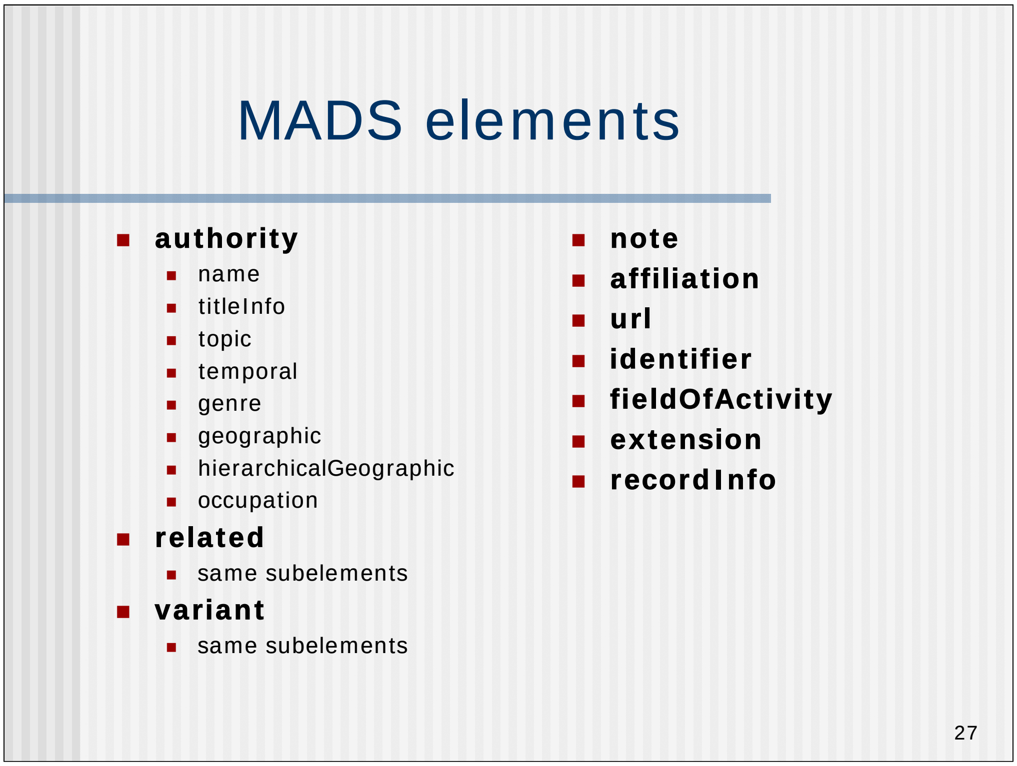
There are many more listed in the MODS implementation registry.

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

26

MADS is a companion to MODS in that it has the same features: a subset of MARC authorities, compatibility with MARC definitions, language based tags in XML.



Some of the MADS data elements are not in MARC authority records in separate fields (although some of the data such as field of activity might be in the 670 Sources found note).

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

28

LC is exploring exposing authority records as MADS.

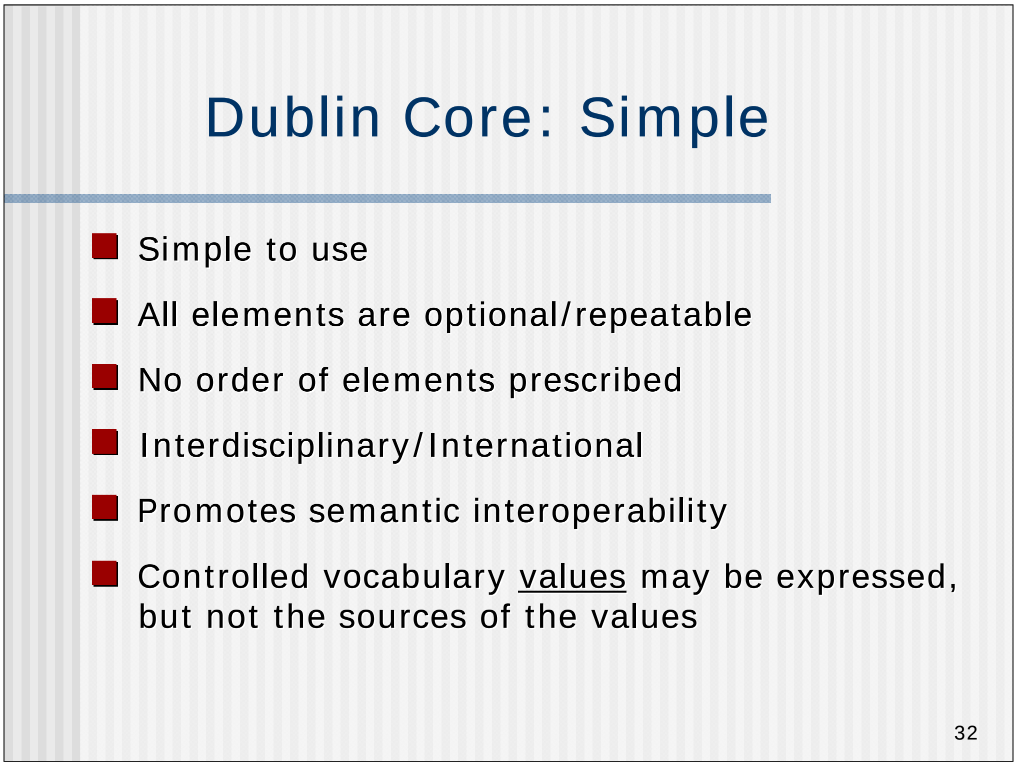
Example: MADS Name Record

```
<mads xsi:schemaLocation="http://www.loc.gov/mads/
    http://www.loc.gov/mads/mads.xsd">
   <authority>
      <name type="personal">
        <namePart>Smith,John</namePart>
        <namePart type="date">1995-</namePart>
     </name>
    </authority>
      <variant type="other">
       <name>
         <namePart>Smith, J</namePart>
       </name>
     </variant>
     <variant type="other">
        <name>
         <namePart>Smith, John J</namePart>
       </name>
      </variant>
   <note type="history">Biographical note about John Smith.</note>
   <affiliation>
      <organization>Lawrence Livermore Laboratory</organization>
      <dateValid>1987</dateValid>
   </affiliation>
</mads>
```

Example: MADS Organization Record

```
<mads xsi:schemaLocation="http://www.loc.gov/mads/
   http://www.loc.gov/mads/mads.xsd">
   <authority>
      <name type="corporate">
        <namePart>Unesco</namePart>
      </name>
    </authority>
   <related type="parentOrg">
      <name>
        <namePart>United Nations</namePart>
      </name>
    </related>
   <variant type="expansion">
      <name>
        <namePart>United Nations Educational, Cultural, and Scientific
        Organization</namePart>
      </name>
   </variant>
</mads>
```

Some digital library projects are beginning to use MADS.



Simplicity: intended to be implemented by catalogers, non-catalogers alike; not meant to replace richer description schema

Interdisciplinary/International: governed by an international group, the Dublin Core Metadata Initiative; annual conferences

Interoperability (general): The ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner. There are three aspects of interoperability: semantic, structural and syntactical.]

Semantic interoperability: Ability to search for digital information across heterogeneous distributed databases whose metadata schemas have been mapped to one another. It is achieved through agreements about content description standards; for example, Dublin Core – MARC; DC-MODS; DC-VRACore. DC is arguably the most frequently “mapped” metadata standard

Controlled vocabulary: In Simple DC you may use controlled vocabulary terms but can’t provide information about where the terms come from, or how the values are encoded (dates, for instance). Qualified DC allows different classification as does simple DC, but qualified DC also allows the cataloger to say where that classification scheme came from. There is more than one technique to use for that (making machine manipulation more difficult).

Qualified DC allows the metadata creator to say what kind of controlled vocabulary is used, and provides techniques to do so in XML.

Dublin Core Elements

Fifteen elements in Simple DC

Title	Creator	Date
Description	Contributor	Coverage
Subject	Publisher	Identifier
Relation	Rights	Format
Source	Coverage	Type

33

By 1996, the 15 elements had been established.

The namespace is where machine (and usually human) readable information can be found.

"Qualified" Dublin Core

- Includes 15 terms of the original DC Metadata Element Set, plus:
- Additional properties and sub-properties
  - Examples: abstract, accessRights, audience, instructionalMethod, rightsHolder, provenance
- Provides:
  - A fuller set of properties with specific requirements for content
  - A namespace that includes all properties
  - Explicit value vocabularies can be specified

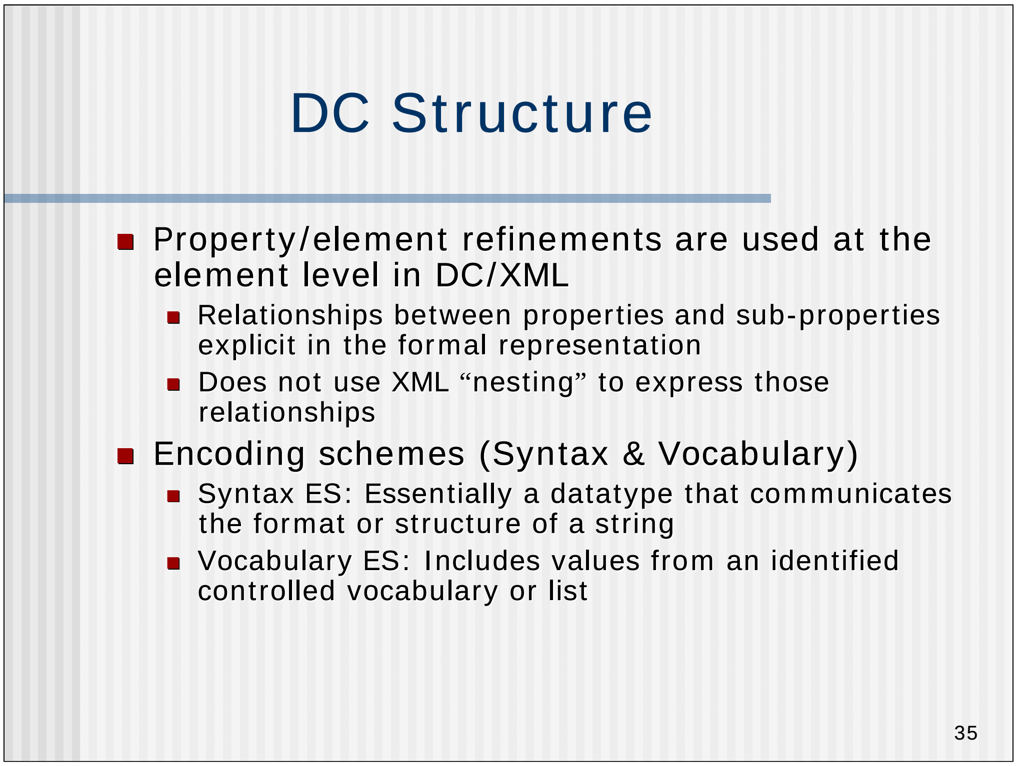
34

Note that "qualifiers" is not currently used, but there's not a generally accepted way to talk about the whole of DC, as distinguished from Simple DC.

Qualified DC has more elements than those listed here see: http://dublincore.org/documents/usageguide/qualifiers.shtml (scroll down to chart)

"Namespace"—used in XML records; establishes uniqueness when attached to a data element so that there is no conflict when two elements have the same name. Formerly, there were separate namespaces for Simple DC and Qualified DC terms. Now, users of Qualified DC only need one namespace. The Simple properties can be used in either the original namespace, without domains and ranges, or in the "dcterms" namespace with domains and ranges.

Explicit vocabularies are generally recommended by a particular domain or community of practice (for example, most library applications use LCSH).

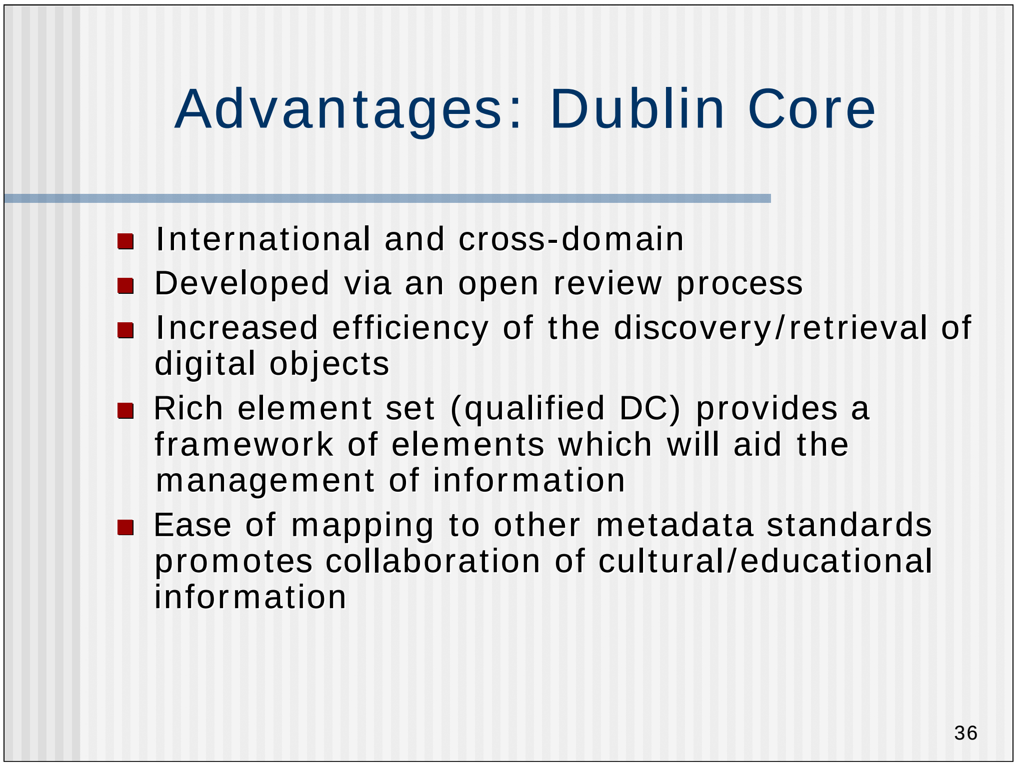


Element refinements (sub-properties) must be related to a “parent” element but the relationship is expressed in the element set declaration, and is not “nested” in XML (for example).

Encoding schemes include syntax encoding schemes and vocabulary encoding schemes used with DC.

Syntax encoding scheme examples include ISO 3166 (codes for names of countries), RFC4646 (codes for languages), W3CDTF (W3C Date and Time Formats Specification).

Vocabulary encoding schemes include DCMI Type Vocabulary, DDC, LCC, LCH, MeSH, etc.



DC is useful especially when disparate metadata formats are used together.

The Dublin Core Metadata Initiative's stated mission is to make it easier to find resources using the Internet through developing metadata standards for resource discovery across domains, defining frameworks for the interoperation of metadata sets, and facilitating the development of community- or domain-specific metadata sets that work within these frameworks.

Use of Qualified DC overcomes many of the limitations of Simple DC, while retaining the same accessible structure. Because DC is non-proprietary and undergoes an open, formal review process, additional refinements are compatible with earlier versions.

DC is also known to be well documented, which makes it easily accessible.

Uses of Dublin Core

- Minimal standard for OAI-PMH
- Core element set in some other schemas
- Switching vocabulary for more complex schemas

37

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) requires at least simple DC in XML; other XML formats may be used in addition to simple DC (e.g. qualified DC, MARCXML, VRA Core).

DC is the “lingua franca” for metadata from diverse communities. It doesn’t “map out” (to richer schemas) as well as it “maps in.”

IEEE-LOM uses simple DC as its core element list

PBCore used simple DC as the basis for its element set

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>text</dc:type>
   <dc:identifier>85753651</dc:identifier>
   <dc:language>it</dc:language>
   <dc:language>en</dc:language>
   <dc:publisher>Nova Music</dc:publisher></metadata>
```

There is no prescribed order of elements in DC. Everything is optional. Notice that the <subject> element has a classification number in it (this is legal, the distinction between classification and topical subject is at the level of the vocabulary). This record also has an identifier, the ISBN (this is a physical object, not a digital one)

DC does not specify a particular syntax.

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>
```

Note the addition of encoding schemes (authorities) for the values in various elements.

Status of DC

- Dublin Core Metadata Element Set version 1.1
  - ISO Standard 15836-2003; ANSI/NISO Standard Z39.85-2007; IETF RFC 5013
- Updated encoding guidelines
  - Proposed recommendation for expressing DC description sets using XML (Sept. 2008)
  - Final recommendation for expressing DC metadata using HTML/XHTML (Aug. 2008)

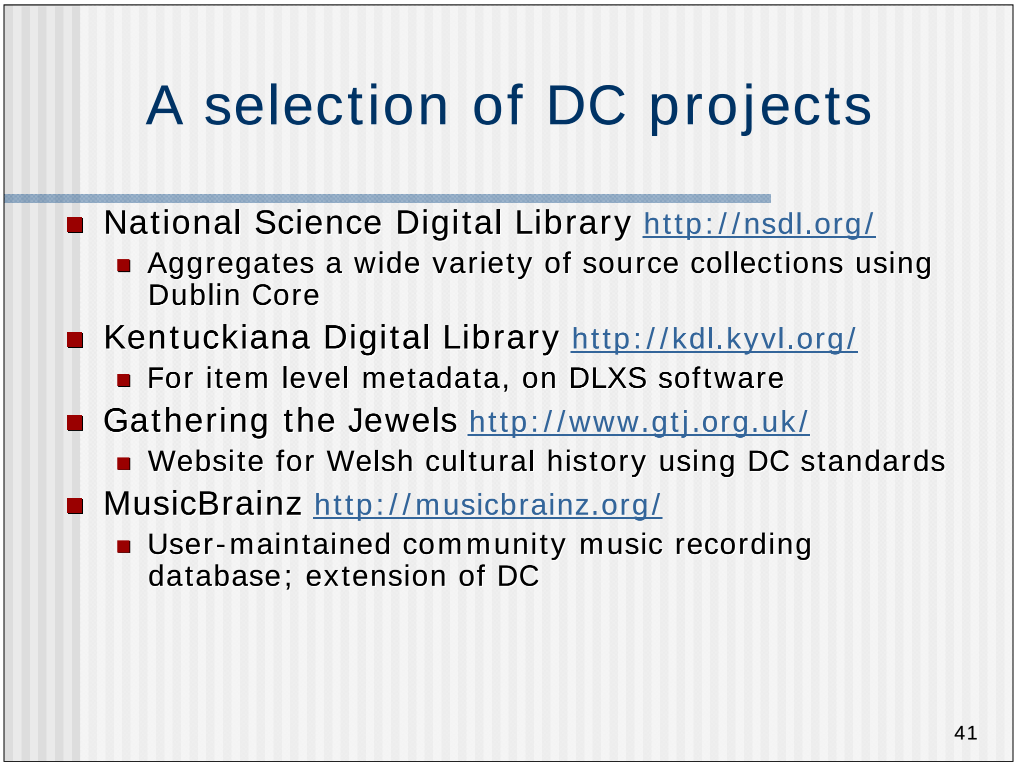
40

Revised ISO 15836 is expected to replace the current 2003 version in 2008. No big changes but improved definitions for many elements

ISO=International Organization for Standards (Geneva), international standards-setting body for business, government and society

ANSI/NISO=NISO, the National Information Standards Organization, is accredited by ANSI, the American National Standards Institute.

IETF=Internet Engineering Task Force, founded 1992, is an open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. RFCs are Requests for Comment. They are the basic publication series for the IETF. Once and RFC is “published” by the RFC editor, it is not modified.



DLXS is the University of Michigan's Digital Library eXtension Service, a licensed repository software platform.

DCMI has a list of registered DC projects at http://dublincore.org/projects/, though it is far from exhaustive. Many digital projects include Dublin Core as an export metadata schema in addition to other primary metadata schemas.

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

42

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

43

EAC=Encoded Archival Context

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

44

Widely used in the archival community. EAD Finding Aids themselves may be described using other metadata schemes (LC has many in its library catalog). There also can be links to other descriptions.

Text Encoding Initiative (TEI)

- Consortium of institutions and research projects which collectively maintains and develops guidelines for the representation of texts in digital form.
- Includes representation of title pages, chapter breaks, tables of contents, as well as poetry, plays, charts, etc.
- The TEI file contains a “header” that holds metadata about the digital file & about the original source.

45

TEI

```
<fileDesc>
    <titleStmt>
        <title type="main">A chronicle of the conquest of
          Granada</title>
        <author>
            <name type="last">Irving</name>
            <name type="first">Washington</name>
            <dateRange from="1783" to="1859">1783-
          1859</dateRange>
        </author>
    </titleStmt>
    <extent>455 kilobytes</extent>
    <publicationStmt>
        <publisher>University of Virginia Library</publisher>
        <pubPlace>Charlottesville, Virginia</pubPlace>
        <date value="2006">2006</date>
    </publicationStmt>
    <availability status="public">
        <p n="copyright">Copyright &copy; 2006 by the Rector
          and Visitors of the University of Virginia</p>
        <p n="access">Publicly accessible</p>
    </availability>
```

<fileDesc> is about the electronic publication (see extent & publicationStmt)

MORE TEI

```
<sourceDesc>
    <titleStmt>
        <title type="main">A chronicle of the conquest of
          Granada</title>
        <author>
        <name type="last">Irving</name>
        <name type="first">Washington</name>
        <dateRange from="1783" to="1859">1783-
          1859</dateRange>
        </author>
    </titleStmt>
    <extent>345 p. ; 21 cm.</extent>
    <publicationStmt>
        <publisher>Carey, Lea &amp; Carey</publisher>
        <pubPlace>Philadelphia</pubPlace>
        <date value="1829">1829</date>
        <idno type="LC call number">DP122 .I7 1829a</idno>
        <idno type="UVa Title Control Number">a1599744</idno>
    </publicationStmt>
</sourceDesc>
```

<sourceDesc> is about the source publication (see extent & publicationStmt). Violates the DC one-to-one rule, but coded explicitly

Selection of TEI projects

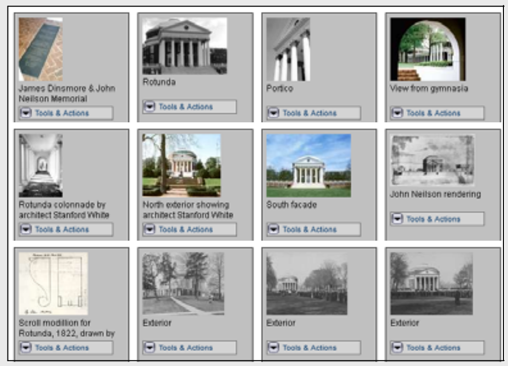
- American Memory (uses a TEI-conformant DTD
  - http://memory.loc.gov/ammem/index.html
- Early Canada Online
  - http://www.canadiana.org/
- Victorian Women Writers Project
  - http://www.indiana.edu/~letrs/vwwp/index.html
- Oxford Text Archive
  - http://ota.ahds.ac.uk/

48

The TEI web site lists many more projects that use TEI.

VRA Core

- Maintained by the Visual Resources Association. Version 4 is currently in beta release.
- A categorical organization for the description of works of visual culture as well as the images that document them.
- Consists of a metadata element set and an initial blueprint for how those elements can be hierarchically structured.



credit: K. Edward Lay

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

50

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)

51

VRA Core

```
<work>
<titleSet>
     <title pref="true” source=“LC NAF”>Rotunda</title>
</titleSet>
<agentSet><agent>
     <name type="personal“ vocab=“LC NAF” refid= “n 79089957”>
               Jefferson, Thomas</name>
     <dates type="life">
             <earliestDate>1743</earliestDate><latestDate>1826</latestDate>
     </dates>
     <role>architect</role>
     <culture>American</culture>
</agent></agentSet>
<agentSet><agent>
     <name type="personal“ vocab=“LC NAF” refid= “n 50020242”>
               White, Stanford</name>
     <dates type="life">
          <earliestDate>1853</earliestDate><latestDate>1906</latestDate>
     </dates>
     <role>architect</role>
     <culture>American</culture>
     <notes>Architect of 1896-1897 renovation</notes>
</agent></agentSet>
```

about the BUILDING (architects)

```
<dateSet>
    <date type="construction">
        <earliestDate>1822</earliestDate><latestDate>1826</latestDate>
    </date>
    <notes>Construction begun October, 1822, completed September, 1826.<notes>
</dateSet>
<dateSet>
    <date type="destruction">
        <earliestDate>1895</earliestDate>
    </date>
    <notes>Burned October 27, 1895.</notes>
</dateSet>
<dateSet>
    <date type="renovation">
        <earliestDate>1896</earliestDate><latestDate>1897</latestDate>
    </date>
    <notes>Rebuilt to designs of Stanford White, 1896-1897.</notes>
</dateSet>
<locationSet><location type="site">
    <name type="geographic" vocab="TGN" refid="2002201">
    Charlottesville, Virginia</name>
</location></locationSet>
</work>
```

about the BUILDING (dates)

More VRA Core

```
<image>
<titleSet>
     <title type="descriptive">general view</title>
</titleSet>
<agentSet><agent>
     <name type="personal“ vocab=“LC NAF”
          refid=“n 82111472”>Lay, K. Edward</name>
     <culture>American</culture>
     <role>photographer</role>
</agent></agentSet>
<dateSet><date type=“creation">
     <earliestDate>1990</earliestDate>
     <latestDate>2000</latestDate>
</date></dateSet>
<locationSet><location type="repository">
     <name type="corporate">University of Virginia Library</name>
     <name type="geographic" vocab="TGN" refid="2002201">
          Charlottesville</name>
</location></locationSet>
<rightsSet>
     <rights type=“credit”>K. Edward Lay</rights>
     <rights type=“access”>Publicly accessible</rights>
</rightsSet>
</image>
```

credit: K. Edward Lay

about the IMAGE (this one is a general view)

More VRA Core

```
<image>
<titleSet>
     <title type="descriptive">View from gymnasia</title>
</titleSet>
<agentSet><agent>
     <name type="personal“ vocab=“LC NAF”
          refid=“n 82111472”>Lay, K. Edward</name>
     <culture>American</culture>
     <role>photographer</role>
</agent></agentSet>
<dateSet><date type=“creation">
     <earliestDate>1995</earliestDate>
     <latestDate>2000</latestDate>
</date></dateSet>
<locationSet><location type="repository">
     <name type="corporate">University of Virginia Library</name>
     <name type="geographic" vocab="TGN" refid="2002201">
          Charlottesville</name>
</location></locationSet>
<rightsSet>
     <rights type=“credit”>K. Edward Lay</rights>
     <rights type=“access”>Publicly accessible</rights>
</rightsSet>
</image>
```

credit: K. Edward Lay

A Selection of VRA Core Projects

- Luna Imaging
  - http://www.lunaimaging.com/index.html
- ARTstor
  - http://www.artstor.org/
- Visual Information Access (VIA), Harvard University Libraries
  - http://via.lib.harvard.edu/via/

56

VIA (Visual Image Access), a union catalog of visual resources at Harvard, draws on early versions of the VRA Core and CDWA.

Learning Object Metadata (LOM)

- 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

57

This standard was developed by IEEE; the IMS Global Learning Consortium is a sort of spin-off (note that neither organization makes their standard or documentary material openly available).

IEEE-LOM uses simple DC as its core element list.

Look at the CanCore implementation, which is really one of the few IEEE-LOM that we can examine. Usage of IEEE-LOM tends to be in closed government sites, for example, Army training manuals.

IEEE-LOM is very rich in its way, but it’s a real commitment.

Uses of IEEE-LOM

- Describe and share information about learning objects individually or as a group
- Export as LOM in XML or RDF
- Most descriptive elements mapped to Dublin Core
- Can be used with the IMS VDEX (Vocabulary Definition Exchange)

58

The VDEX specification is a grammar for controlled vocabularies; a mark-up language for CVs.

A Selection of IEEE-LOM Projects

- CanCore
  - http://www.cancore.ca/
- LearnAlberta.ca
  - http://www.learnalberta.ca/
  - Grades K-12
- Learning Object Repository Network
  - http://lorn.flexiblelearning.net.au/Home.aspx

59

What is 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.

60

Briefly.

It’s hard to map a second hierarchical version. ONIX includes a lot of marketing text and information helpful to shipping, e.g., taxes in different jurisdictions, height and weight. Some information is very useful to libraries: book jacket images, for instance.

Advantages of ONIX

- Provides publisher information in a widely used standard format
- Promotes exchange of information with publishers, vendors, book sellers, libraries
- “Value-added” information (ex., book jacket images, reviews) benefits book sellers (online commercial sites) and libraries (online catalogs)
- More [information], faster [transmission], cheaper? better?

61

ONIX has streamlined ordering, allowing faster and more reliable electronic transmissions. It also allows the exchange of bibliographic information from the publisher-vendor to libraries, though sometimes the content does not conform to the grammatical format traditionally used by libraries. (Capitalization, punctuation vary)

A selection of ONIX projects

- http://www.editeur.org/onix.html
- ONIX Administrators
  - EDItEUR (European & international)
  - Book Industry Communication (BIC) (European and international)
  - Book Industry Study Group, Inc. (BISG) (U.S.)
- Amazon.com
- Association of American Publishers
- Baker & Taylor
- Barnes & Noble
- Google
- McGraw-Hill Companies

62

PBCore

- Public Broadcasting Core element set
  - http://www.pbcore.org/
  - Built on Dublin Core (but does not comply with the Abstract Model)
- Provides a shared descriptive language for public broadcasters
  - Used for television, radio, Web activities

63

PBCore Elements

- 53 elements arranged in 15 containers and 3 sub-containers
- Four classes:
  - Intellectual Content (title, subject, description, audienceLevel ...)
  - Intellectual Property (creator, contributor, publisher, rightsSummary)
  - Instantiation (dateCreated, formatFileSize, formatDuration, formatTracks, language)
  - Extensions

64

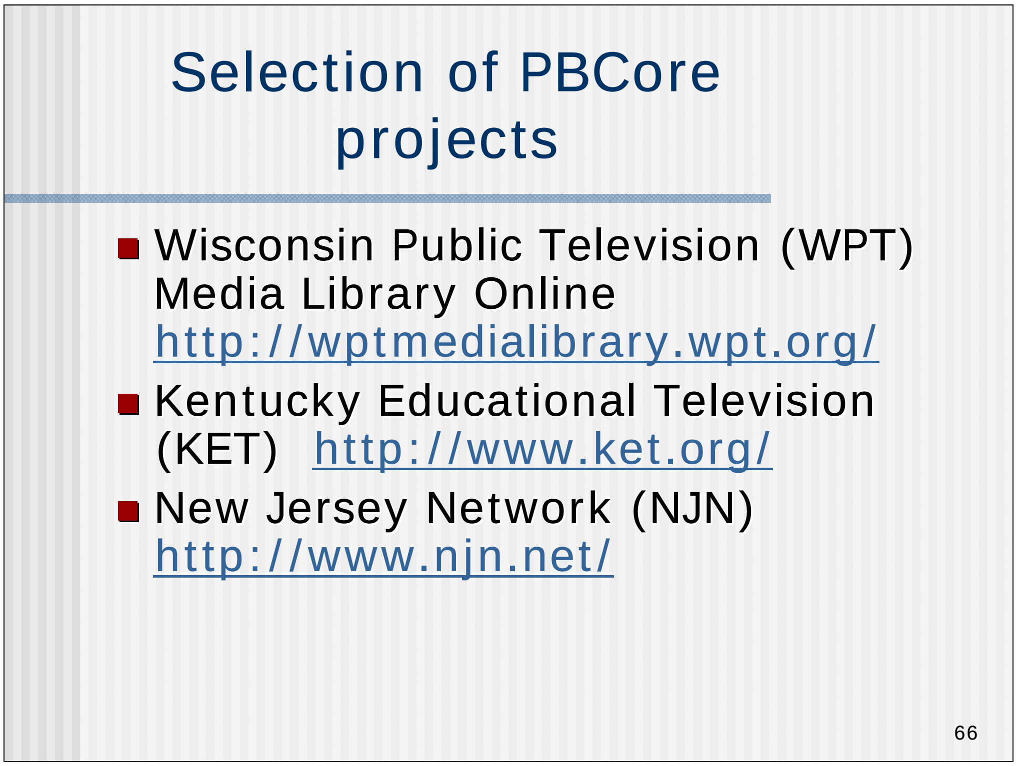
Note that the first four classes similar to those DC used until recently to categorize its elements.

Uses of PBCore

- Shared descriptive language for public broadcasters
- Useful for both public search and viewing, and internal asset management
- Facilitates production collaborations
- Ability to parse programs into short segments for Web distribution, niche community needs

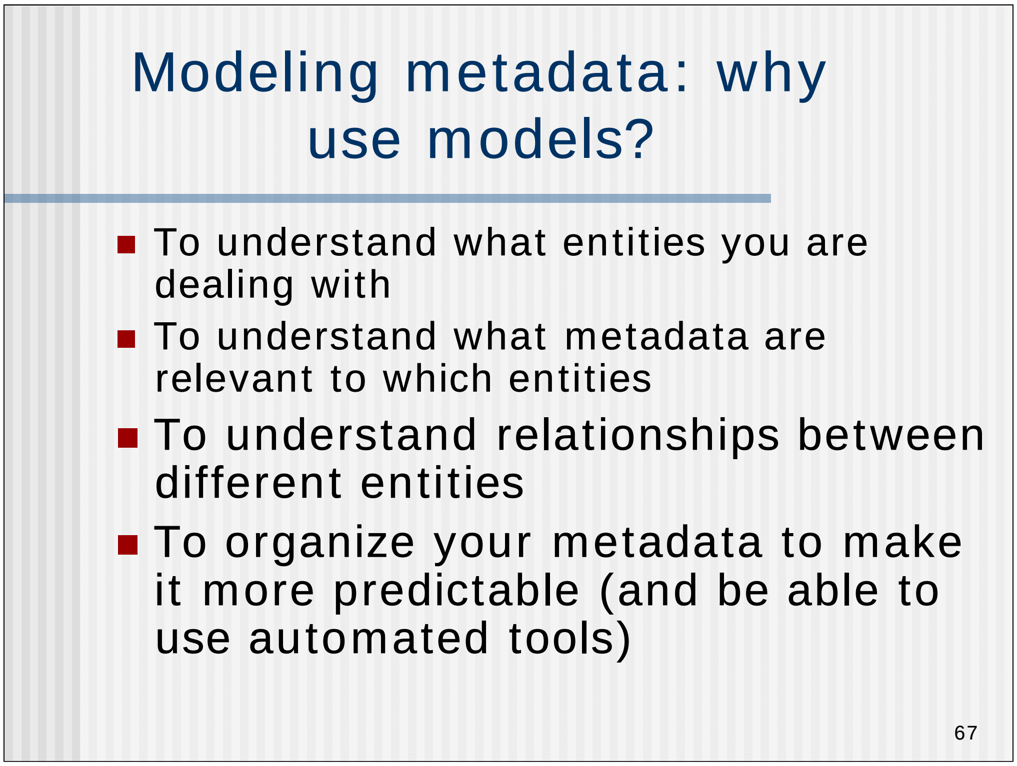
65

Granular manipulation and interoperability of content are required in order to allow re-use of media content.



Most public television stations/networks do not have the infrastructure to support a public, online library of moving images. WPT was the first metadata management system to use the PBCore metadata dictionary.

PBCore is commonly used as the internal standard to describe and track a station's or network's moving image assets. NJN is beginning to make some of its moving images available through the New Jersey Digital Highway (a statewide repository of cultural and historical objects) and NJVid (a statewide video commons for the state's K-20 institutions, museums and libraries).



Modeling standards, like FRBR and FRAD, aim to facilitate the automated exchange of data and services between remote computer systems by defining a model for the data or the services.

Here are some reasons why models might be important in developing metadata standards.

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

68

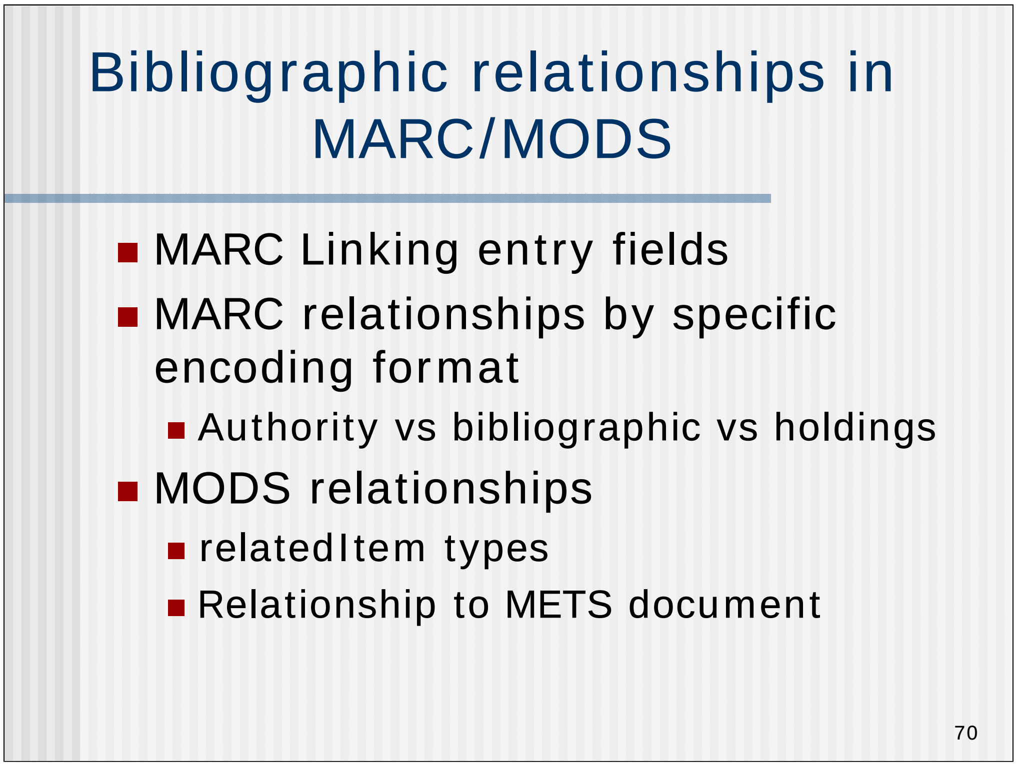
Models may be widely variant in granularity and in use. Some may model information, while others functionality. Some may be for specific kinds of metadata, others for specific kinds of resources. Very often models are developed after the development of the standard itself.

Bibliographic relationships (pre-FRBR)

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

69

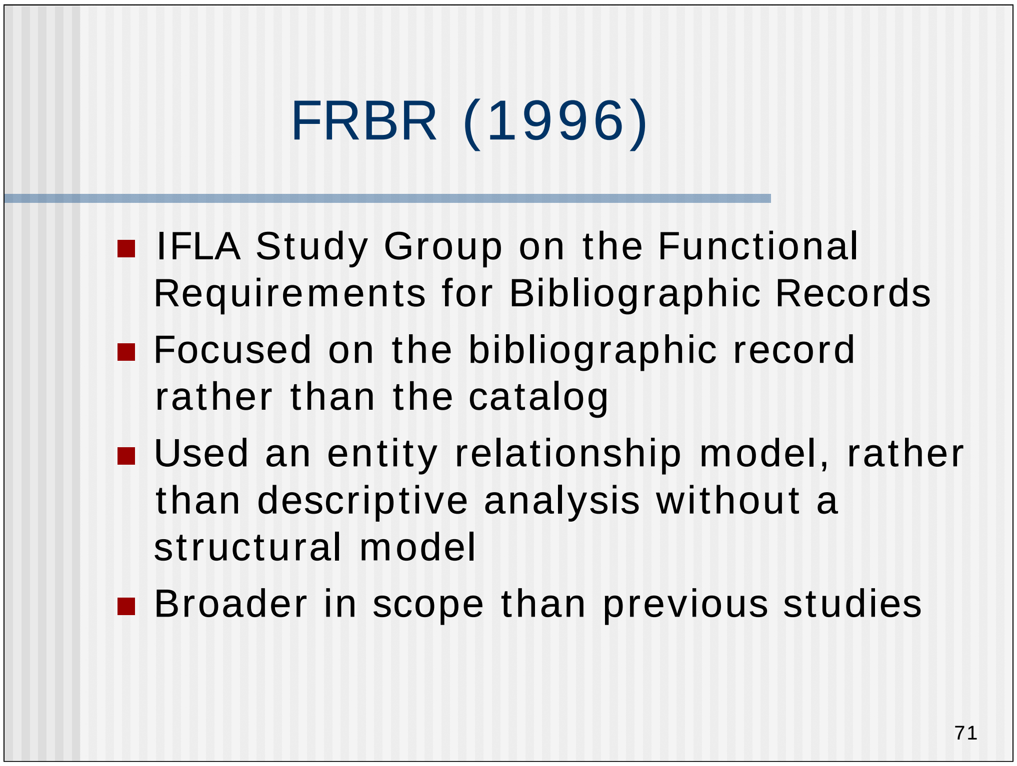
Although dated, this taxonomy is useful for detailing types of relationships between bibliographic resources.



MARC linking entry fields (although developed initially for serials, but later made applicable to all types) are defined for many of the relationships in Tillett's taxonomy.

There are also relationships between entities described in the different MARC formats (somewhat related to FRBR entities).

MODS relationships are rich and detailed in related item types. They have a close relationship with the structural metadata detailed in a METS document (we will cover this in session 4).



This is worth looking at carefully when considering broader implications.

FRBR is more than the bibliographic entities (work-expression-manifestation-item) that are usually discussed in the library world. It asks important questions:

What are the relationships between entities and people (agents)?

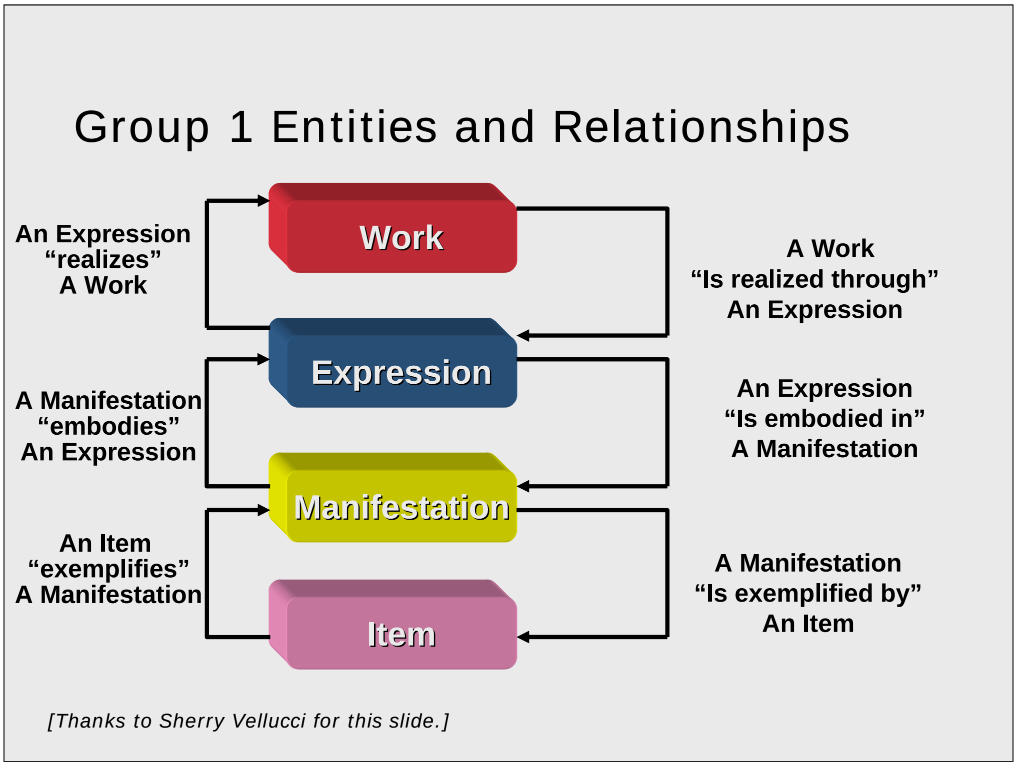
What are the implications for how a bibliographic system should work?

FRBR Entities

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

72

Three classes of FRBR entities.



The Group 1 entities are the Work, the Expression, the Manifestation and the Item

The diagram shows the relationships between two adjacent entities

DC Abstract Model

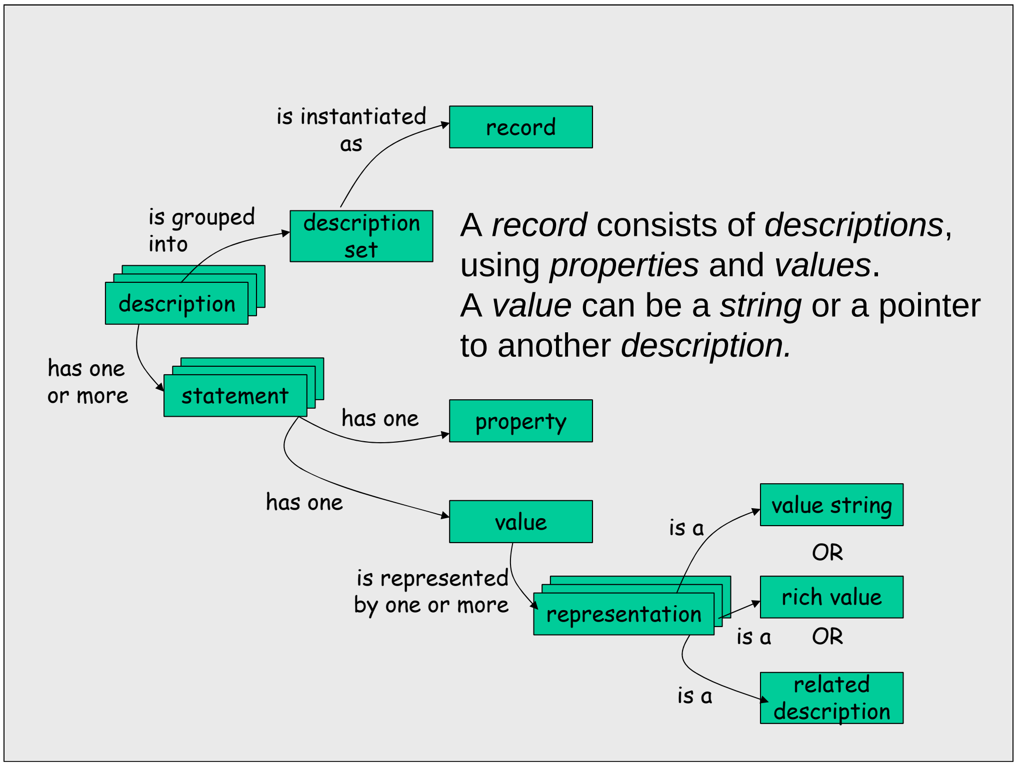
- Reaffirms the One-to-One Principle
- Defines ‘statement’ as the atomic level
- Distinguishes between “description” and “description set”:
  - **Description**: “A description is made up of one or more statements about one, and only one, resource.”
  - **Description Set**: “A description set is a set of one or more descriptions about one or more resources.”
- RDA vocabularies being developed to use the DC Abstract Model

74

The model defines the metadata statement as the most granular level– the atomic level.

Groups metadata statements into description and description set.

The DC abstract model clusters related entities.



The metadata statement is core.

Example of a metadata statement: 245 1_ Moby Dick

A value can be a string or something else (e.g., an image). An image is a rich value.

The property of the statement in our example would be the title.

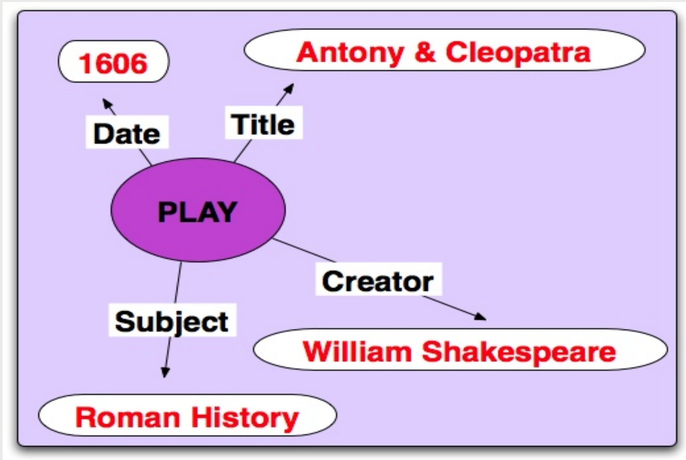
A statement in that description should describe one bibliographic entity (one-to-one-rule).

A series of descriptions can be grouped into one or more description sets.

That description set gets grouped into a record.

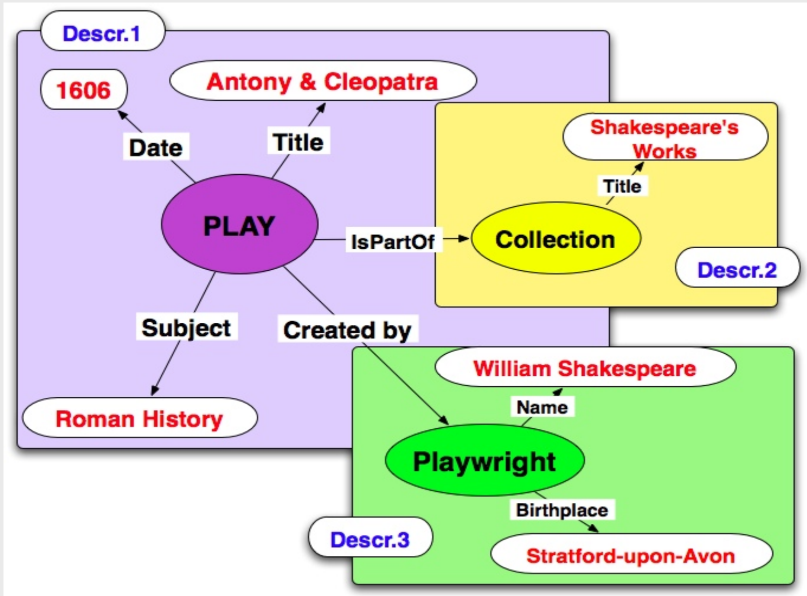
One way to keep from violating the model is to understand the roles of (and not conflate) attributes and values.

Basic model: Resource with properties

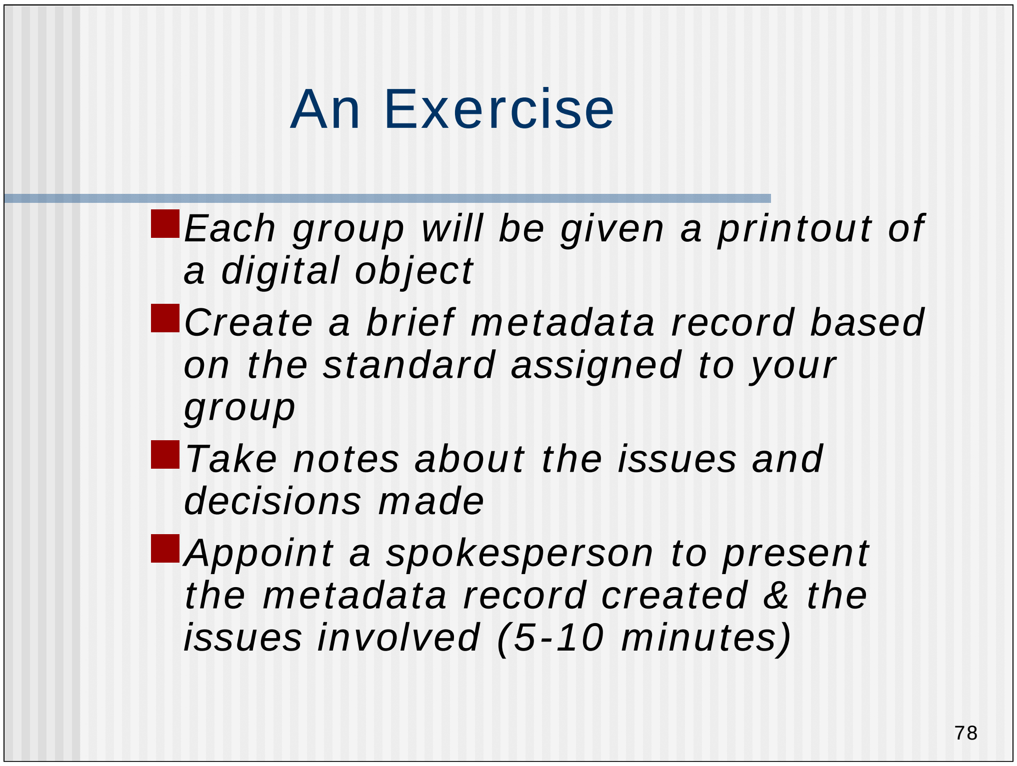


A Play has the title “Antony and Cleopatra,” was written in 1606 by William Shakespeare, and is about “Roman history”

... related to other Resources



Play has a title, an author. The author has a name, birthplace, etc.



LC is providing 2 objects with information for trainees and answer sheets for instructors to hand out after the exercise is completed.

This may be done as an individual or group exercise.

What controlled vocabularies would you consider? Are all fields necessary? What are some differences in approach between the different descriptive metadata standards?

More information is in the Instructors' Notes.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Exercise for Session 2: Descriptive metadata
Student instructions and Template

Example 1:
America's pinch hit march
Sheet music; 3 pages (2 pages music with cover)
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200033287/default.html
(images are in the packet)

Example 2:
52nd Street, New York, N.Y.
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.gottlieb.02771/default.html
(image is in the packet)

Fill in the following metadata elements in the metadata scheme assigned:

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

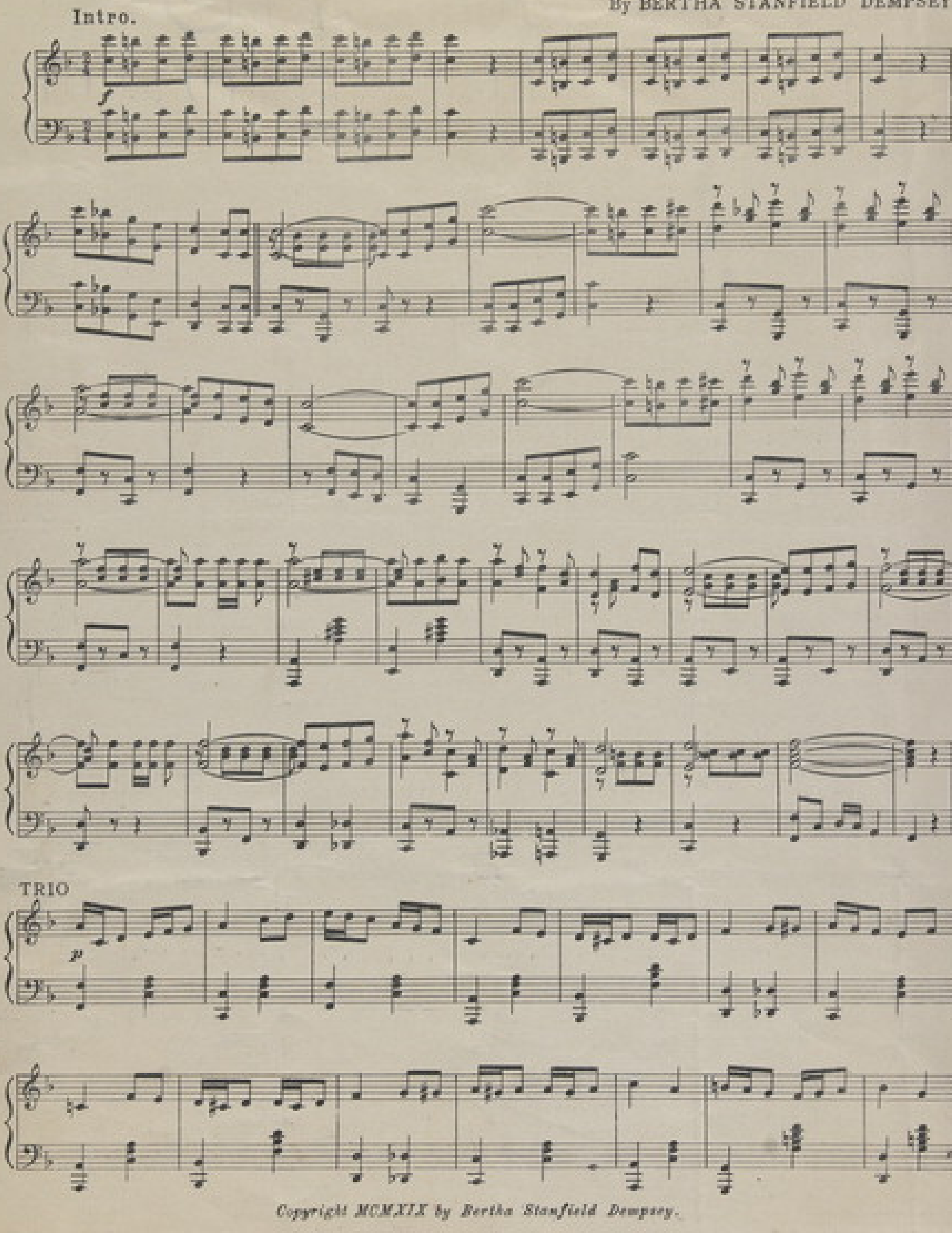
THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Music Division, Reserve Storage.
M1978
.S7D

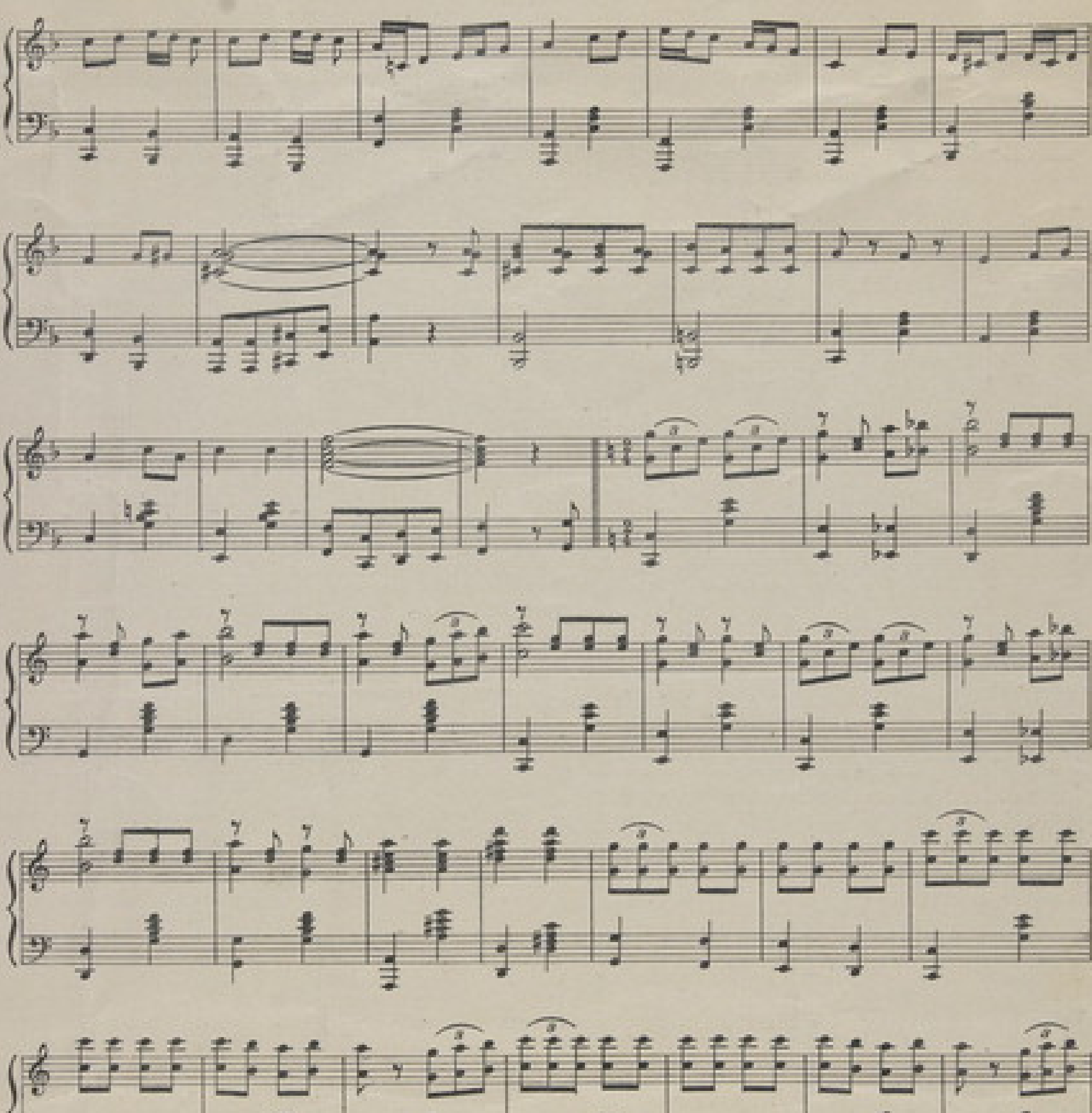
America's Pinch Hit March

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

By BERTHA STANFIELD DEMPSEY



Copyright MCMXIX by Bertha Stanfield Dempsey.



THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

[52nd Street, New York, N.Y., ca. 1948] / William P. Gottlieb [photograph]

default view | **640** | 1024

default view | **640** | 1024

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Template

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

Template

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

Template

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

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]
- ANSI/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/z39-85-2007/
[ISO15836]	http://www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on
[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/

Errata
2008-06-09. Updated URIs for ANSI/NISO Z39.85 and ISO 15836.

Copyright © 1995-2008 DCMI. All Rights Reserved.

Dublin Core Metadata Initiative
Making it easier to find information

DCMI Type Vocabulary

Title:	DCMI Type Vocabulary
Creator:	DCMI Usage Board
Identifier:	http://dublincore.org/documents/2008/01/14/dcmi-type-vocabulary/
Date Issued:	2008-01-14
Latest Version:	http://dublincore.org/documents/dcmi-type-vocabulary/
Replaces:	http://dublincore.org/documents/2006/08/28/dcmi-type-vocabulary/
Replaced By:	Not applicable
Translations:	http://dublincore.org/resources/translations/
Document Status:	This is a DCMI Recommendation.
Description:	The DCMI Type Vocabulary provides a general, cross-domain list of approved terms that may be used as values for the Resource Type element to identify the genre of a resource. The terms documented here are also included in the more comprehensive document "DCMI Metadata Terms" at http://dublincore.org/documents/dcmi-terms/.

Term Name: Collection	
URI:	http://purl.org/dc/dcmitype/Collection
Label:	Collection
Definition:	An aggregation of resources.
Comment:	A collection is described as a group; its parts may also be separately described.
Type of Term:	Class

Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#Collection-003
Term Name: Dataset	
URI:	http://purl.org/dc/dcmitype/Dataset
Label:	Dataset
Definition:	Data encoded in a defined structure.
Comment:	Examples include lists, tables, and databases. A dataset may be useful for direct machine processing.
Type of Term:	Class
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#Dataset-003
Term Name: Event	
URI:	http://purl.org/dc/dcmitype/Event
Label:	Event
Definition:	A non-persistent, time-based occurrence.
Comment:	Metadata for an event provides descriptive information that is the basis for discovery of the purpose, location, duration, and responsible agents associated with an event. Examples include an exhibition, webcast, conference, workshop, open day, performance, battle, trial, wedding, tea party, conflagration.
Type of Term:	Class
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#Event-003
Term Name: Image	

URI:	http://purl.org/dc/dcmitype/Image
Label:	Image
Definition:	A visual representation other than text.
Comment:	Examples include images and photographs of physical objects, paintings, prints, drawings, other images and graphics, animations and moving pictures, film, diagrams, maps, musical notation. Note that Image may include both electronic and physical representations.
Type of Term:	Class
Broader Than:	http://purl.org/dc/dcmitype/StillImage
Broader Than:	http://purl.org/dc/dcmitype/MovingImage
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#Image-004
Term Name: InteractiveResource	
URI:	http://purl.org/dc/dcmitype/InteractiveResource
Label:	Interactive Resource
Definition:	A resource requiring interaction from the user to be understood, executed, or experienced.
Comment:	Examples include forms on Web pages, applets, multimedia learning objects, chat services, or virtual reality environments.
Type of Term:	Class
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#InteractiveResource-003
Term Name: MovingImage	

URI:	http://purl.org/dc/dcmitype/MovingImage
Label:	Moving Image
Definition:	A series of visual representations imparting an impression of motion when shown in succession.
Comment:	Examples include animations, movies, television programs, videos, zoetropes, or visual output from a simulation. Instances of the type Moving Image must also be describable as instances of the broader type Image.
Type of Term:	Class
Narrower Than:	http://purl.org/dc/dcmitype/Image
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#MovingImage-003
Term Name: PhysicalObject	
URI:	http://purl.org/dc/dcmitype/PhysicalObject
Label:	Physical Object
Definition:	An inanimate, three-dimensional object or substance.
Comment:	Note that digital representations of, or surrogates for, these objects should use Image, Text or one of the other types.
Type of Term:	Class
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#PhysicalObject-003
Term Name: Service	
URI:	http://purl.org/dc/dcmitype/Service
Label:	Service

Definition:	A system that provides one or more functions.
Comment:	Examples include a photocopying service, a banking service, an authentication service, interlibrary loans, a Z39.50 or Web server.
Type of Term:	Class
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#Service-003
Term Name: Software	
URI:	http://purl.org/dc/dcmitype/Software
Label:	Software
Definition:	A computer program in source or compiled form.
Comment:	Examples include a C source file, MS-Windows .exe executable, or Perl script.
Type of Term:	Class
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#Software-003
Term Name: Sound	
URI:	http://purl.org/dc/dcmitype/Sound
Label:	Sound
Definition:	A resource primarily intended to be heard.
Comment:	Examples include a music playback file format, an audio compact disc, and recorded speech or sounds.
Type of Term:	Class
Member Of:	http://purl.org/dc/terms/DCMIType

Version:	http://dublincore.org/usage/terms/history/#Sound-003
Term Name: StillImage	
URI:	http://purl.org/dc/dcmitype/StillImage
Label:	Still Image
Definition:	A static visual representation.
Comment:	Examples include paintings, drawings, graphic designs, plans and maps. Recommended best practice is to assign the type Text to images of textual materials. Instances of the type Still Image must also be describable as instances of the broader type Image.
Type of Term:	Class
Narrower Than:	http://purl.org/dc/dcmitype/Image
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#StillImage-003
Term Name: Text	
URI:	http://purl.org/dc/dcmitype/Text
Label:	Text
Definition:	A resource consisting primarily of words for reading.
Comment:	Examples include books, letters, dissertations, poems, newspapers, articles, archives of mailing lists. Note that facsimiles or images of texts are still of the genre Text.
Type of Term:	Class
Member Of:	http://purl.org/dc/terms/DCMIType
Version:	http://dublincore.org/usage/terms/history/#Text-003

Copyright © 1995-2008 DCMI. All Rights Reserved.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Metadata Object Description Schema (MODS) Official Web Site

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 sources where applicable. It is an "outline" of the schema. Items highlighted in red indicate changes made to MODS in Version 3.3.

All top-level elements and all attributes are optional, but you must have at least one element. Subelements are optional, although in some cases you may not have empty containers. Attributes are not in a mandated sequence and not repeatable (per XML rules). "Ordered" below means the subelements must occur in the order 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 to documents 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 Guidelines.

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/relatorsource.html)
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 (approximate, inferred, questionable)

dateCreated

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

dateCaptured

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

dateValid

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

dateModified

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

copyrightDate

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

dateOther

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approximate, inferred, 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: material, 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 implemented note types, 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 (approximate, inferred, 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, otherFormat, 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, music publisher, 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; transliteration; xlink

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, questionable)
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);
lang; xml:lang; script; transliteration

recordCreationDate

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

recordChangeDate

Attributes: encoding (w3cdtf, iso8601, marc); point (start, end); keyDate (yes); qualifier (approximate, inferred, 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]

Questions and comments:
Contact Us (September 4, 2008)

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

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**

Appendix 1: Related Data Standards (when online versions are available they are listed below)

Data element sets:

Categories for the Description of Works of Art
http://www .getty .edu/research/cond ucting research/standards/cdwa/

Data content:

Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images/ editors, Murtha Baca ... [et al.l. on behalf of the Visual Resources Association. Chicago: American Library Association, 2006. See also http://vraweb.org/ccoweb/cco/index.html

Anglo-American cataloguing rules / prepared under the direction of the Joint Steering Committee for Revision of AACR, a committee of the American Library Association ... let al.l. 2nd ed., 2002 revision. Ottawa: Canadian Library Association; Chicago: American Library Association, 2002- (AACR2)

Data values:

Library of Congress Subject Headings (LCSH)

Library of Congress Thesaurus for Graphic Materials (LCTGM - Parts I and II)
http://www.loc.gov/rr/prinUtgm1/
http://www .loc. gOY/rr/prinUtgm2/

Getty Art and Architecture Thesaurus (AA T)
http://www.getty.edu/research/conductingresearch/vocabularies/aaUindex.html

Getty Thesaurus of Geographic Names (TGN)
http://www.getty.edu/research/conductingresearch/vocabularies/tgn/index.html

Getty Union List of Artist Names (ULAN)
http://www.getty.edu/research/conductingresearch/vocabularies/ulan/index.html

Getty Editorial Guidelines for ULAN: Appendix G: Nationalities and Places
http://www.getty.edu/research/conducting research/vocabularies/guidelines/ulan 4 7 ap pendix 9 nationality place.pdf

Appendix 2: Recommended XML introductory resources

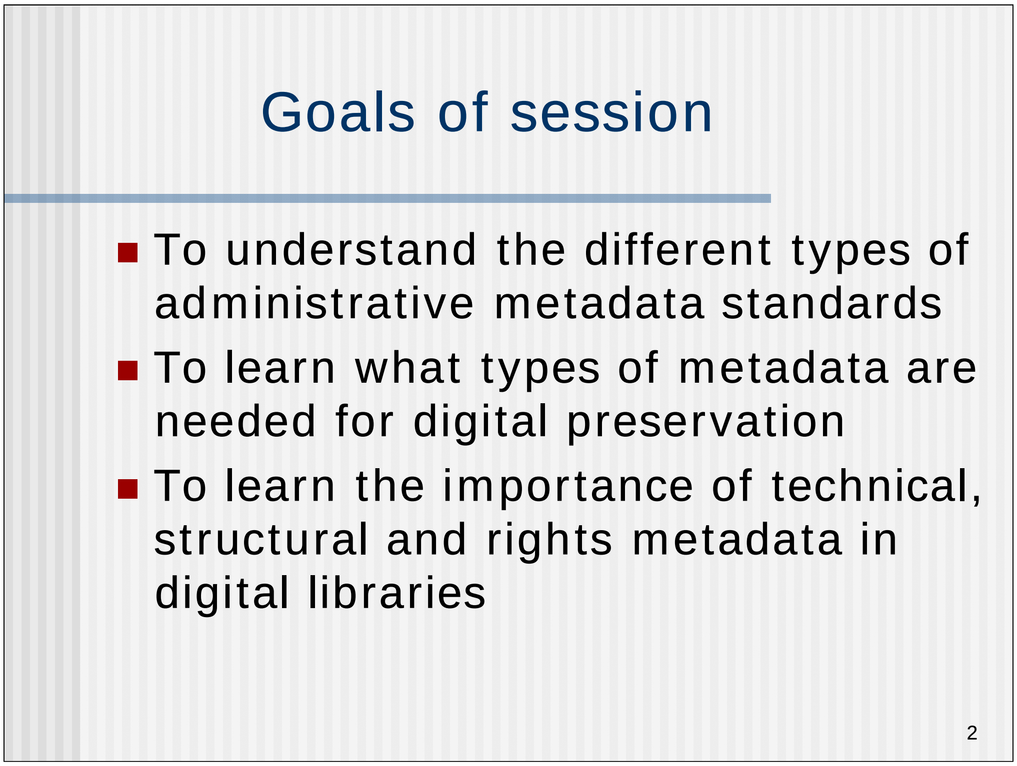
XML tutorial put out by w3Schoois
http://www.w3schools.com/xml/defau lt.asp

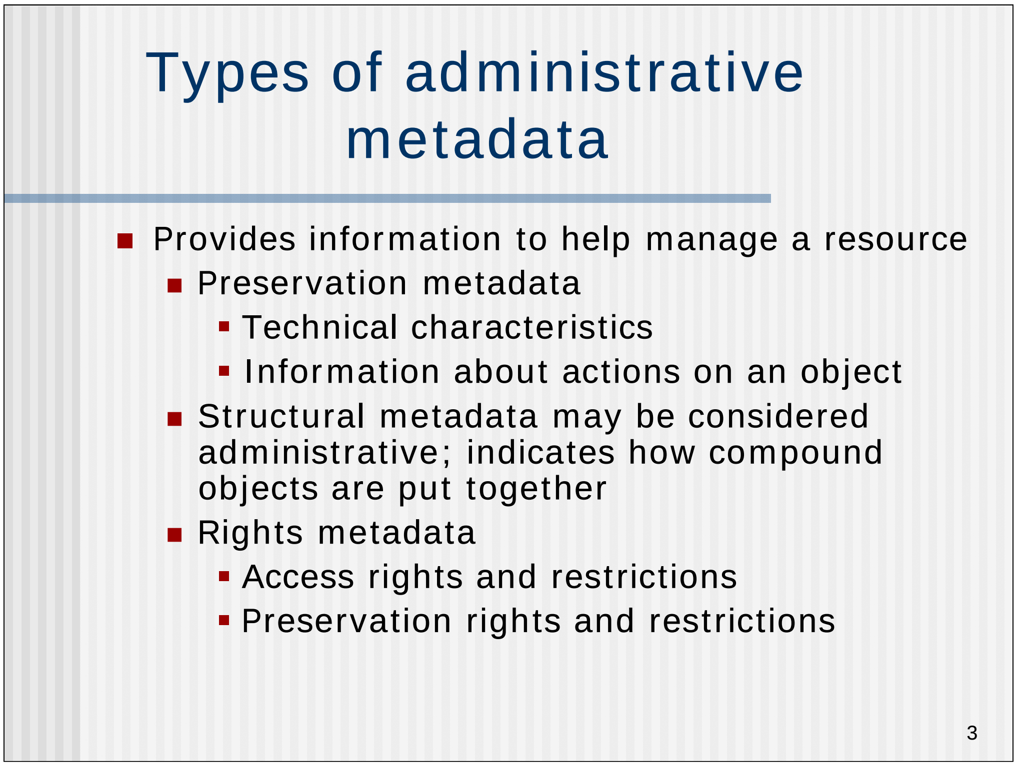
Eric Lease Morgan's Getting Started with XML
http://www.infomotions.com/musings/getting-started/

Help files within software programs- For example, in Access search for "XML for the uninitiated"

Gilmour, Rom. XML: a Guide for Librarians. L1TA Guide #11. American Library Association, 2003.


3. Technical and administrative metadata standards
Metadata Standards and Applications Workshop


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



Administrative metadata gives information about managing the object such as how and why it was created, file format information, actions performed on objects, and information about who can access it. These are sometimes divided into subsets, such as “preservation metadata”, “structural metadata” and “rights metadata” There can be some overlap with elements in descriptive metadata. For instance something like extent or file size is important for the descriptive task of identifying the resource appropriate for the intended use, while it is also important for managing the file as a physical object.

PREMIS: introduction

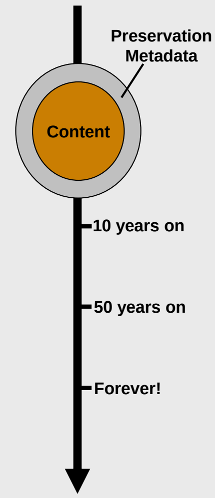
- Preservation metadata that includes subcategories:
  - Technical metadata
  - Relationships (structural and derivative)
  - Digital provenance (what actions performed on objects)
  - Rights

4

PREMIS stands for PREservation Metadata: Implementation Strategies Working Group, which developed a data dictionary of elements needed to support the preservation of digital objects. Preservation metadata falls into various categories, each of which also support other uses (such as access).

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 it?*
- Technical Environment:
  - *What is needed to render and use it?*
- Rights Management:
  - *What IPR must be observed?*



➢ **Makes digital objects self-documenting across time**

There are a number of questions that need to be answered to ensure preservation of digital objects over time. We all know that files can get corrupted and unreadable and it is important to be able to track who has had custody of an object and all actions performed on an object. What preservation strategies have been applied needs to be recorded as well as what software or hardware is needed to use it and whether there are any restrictions on preservation activities (in addition to access). If metadata is associated with an object it can become self-documenting over time.

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

6

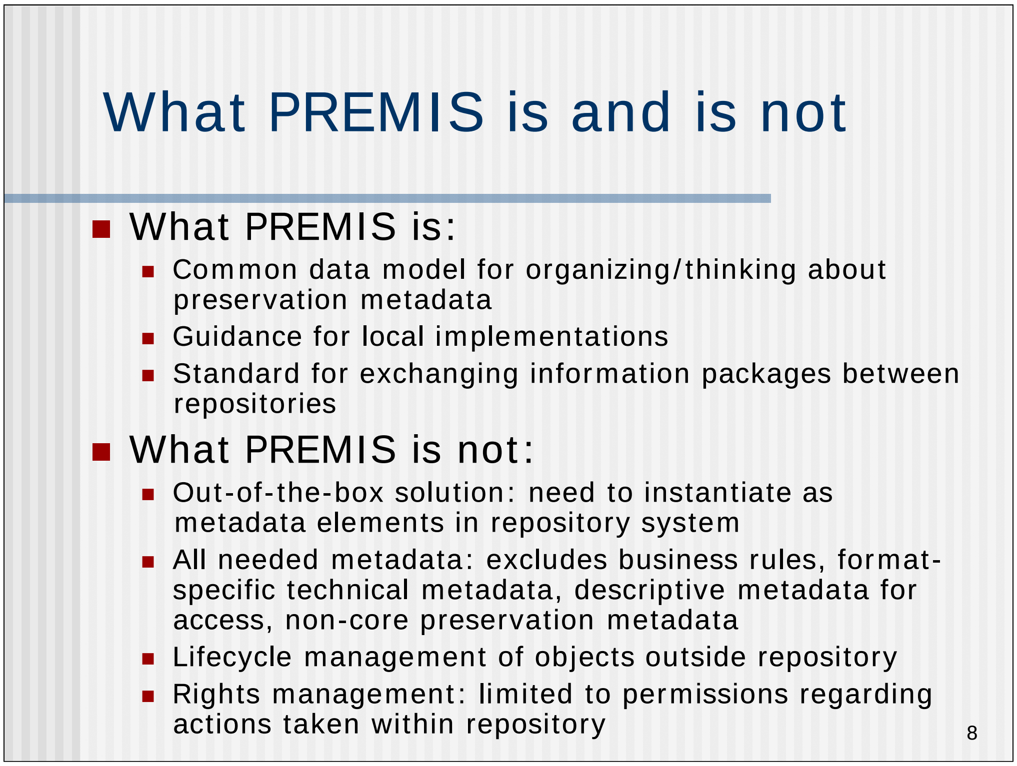
The PREMIS Data Dictionary has become the standard for core metadata needed for preservation activities.

The PREMIS Data Dictionary assumes that implementation decisions need to be made and does not dictate a syntax, although an XML schema is provided.

Descriptive metadata only includes an identifier, but otherwise is out of scope, although its importance is recognized.

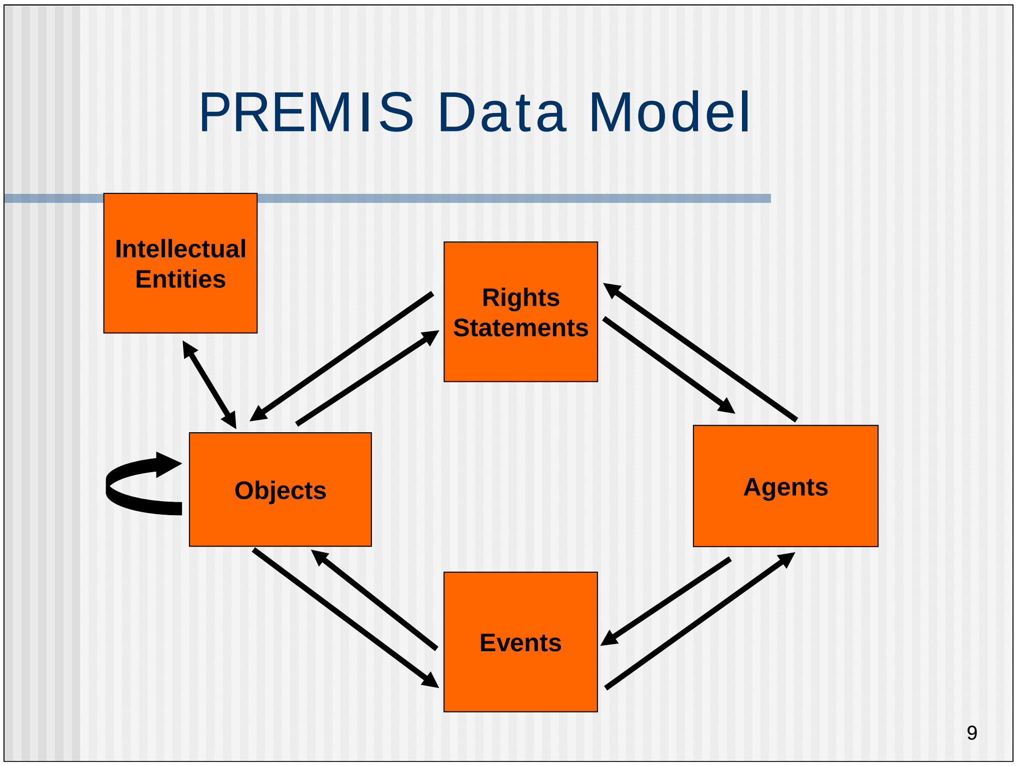
Technical metadata that is included is only that which applies to all format types, not specific file format types (e.g. image, video, etc.)

Business rules are policy decisions that a repository makes about preservation strategies that may or may not be relevant to another institution. Some examples are retention periods or risk assessment.

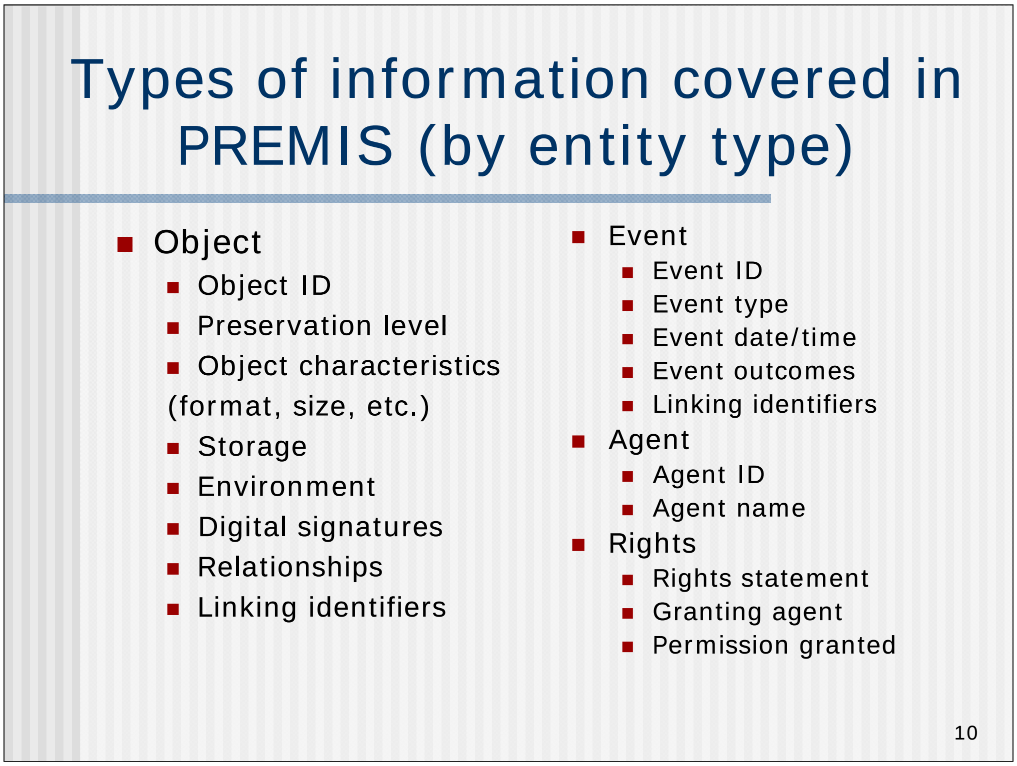


PREMIS may be used in various ways. There is an emphasis on its usefulness as an exchange format between repositories, where a standard is clearly needed. Internally, it may serve as a checklist for core metadata included in the repository.

It needs to be used in conjunction with other metadata and is only one part of the infrastructure needed for digital preservation.



These are the relationships in the PREMIS data model. Although descriptive metadata is out of scope, objects can point to descriptions of intellectual entities or the entity itself, in either direction. From the data model you see that agents (people, organizations or software) can only act on objects through events or rights, not directly. Objects can have relationships to other entities in the data model or to each other.



These are the categories of information covered by PREMIS. The data dictionary is organized by entity type and lists elements that are properties of the given entity.

Objects include technical metadata, much of which is included under “objectCharacteristics”. In addition to format and size, it also includes things like fixity (commonly known as “checksums”, which is used for managing the authenticity of objects), and any passwords or encryption that limits the use of an object. Environment details the software and hardware that is needed to render or use the object. Relationships are documented and are considered either structural (to understand the structure of complex digital objects if you need to put them back together) or derivative (if you make a new copy from an existing object, for example if you make a service copy from a master copy).

Events detail any actions taken on an object. The type of event is in eventType, which includes actions such as capture, compression, validation, replication, migration, etc. These are actions that are performed on objects as part of a preservation strategy. Events are extremely important for preservation activities, because it is necessary to track actions performed on files for preserving into the future.

Agents (people, organizations, software) are not described in great detail, only enough for identification (it is assumed that more information is available elsewhere).

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

Membership of Editorial Committee: Library of Congress, OCLC, FCLA, National Archives of Scotland, British Library, National Library of Australia, U. of Goettingen, LANL, Ex Libris, Library Archives Canada

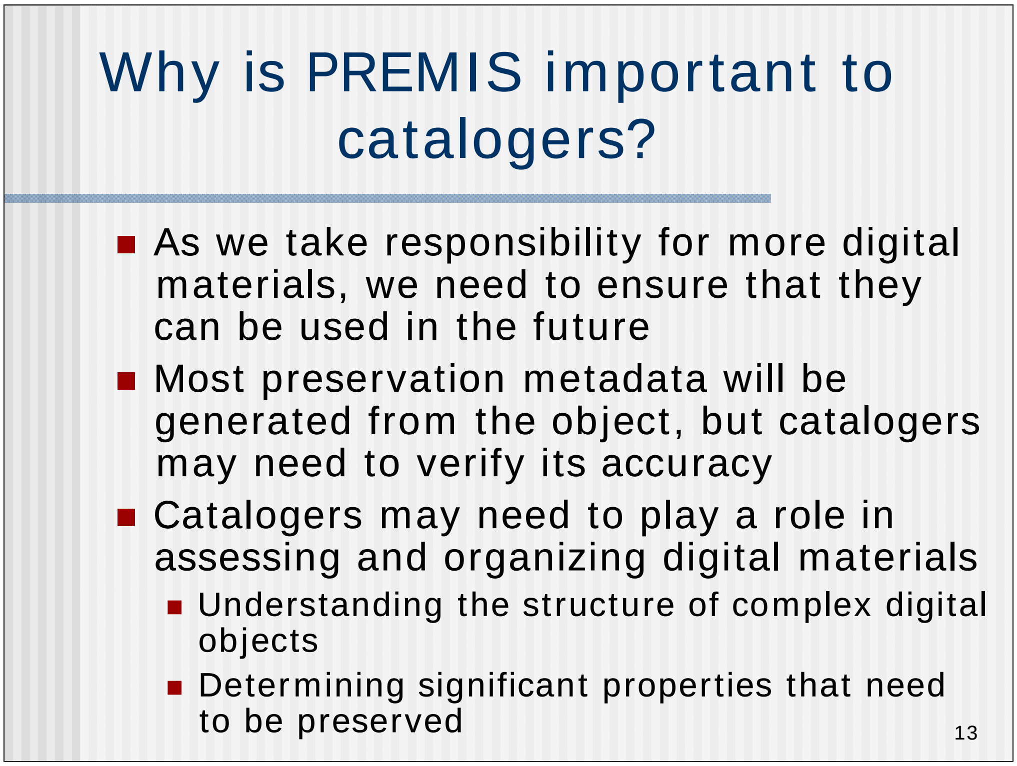
The PIG list is open for potential or current implementers.

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 have been developed
- PREMIS tutorials
  - One or one and a half day tutorials have been given in several locations: Glasgow, Boston, Stockholm, Albuquerque, Washington, San Diego, Berlin
  - Training materials available from LC

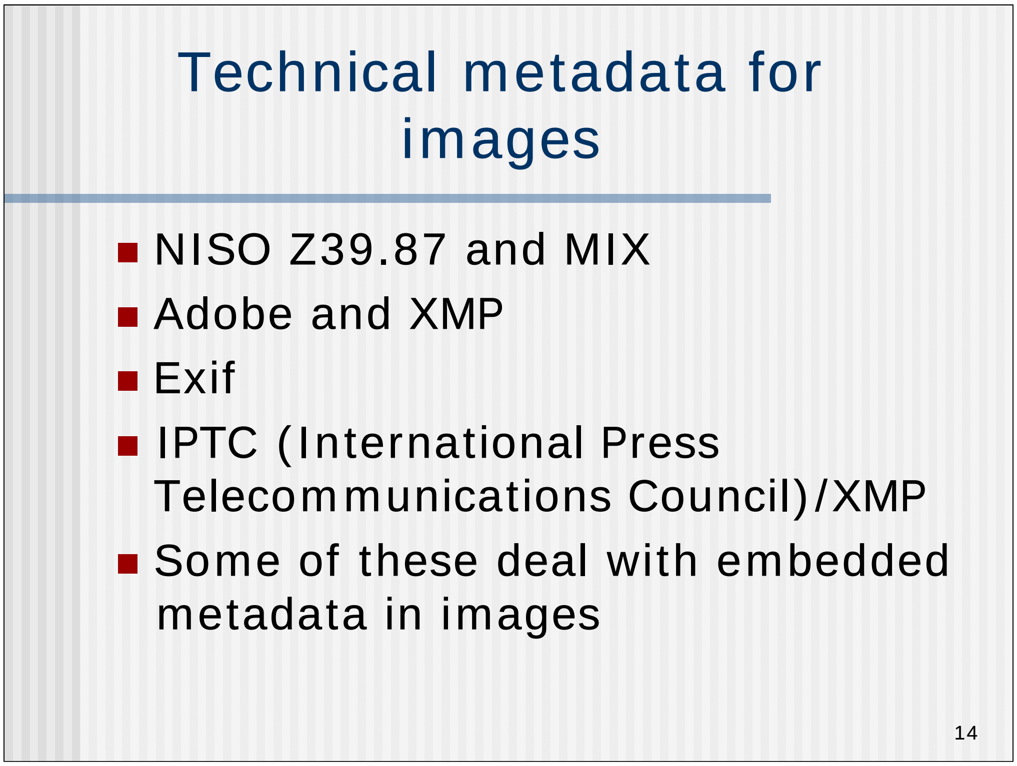
12

Version 2.0 was released the end of March 2008. We will talk about how PREMIS is used with METS later.



Catalogers have always done some assessment of material. Also, there are data elements that serve a function both for the key user tasks of bibliographic (descriptive) records and for preservation purposes.

Although much of the PREMIS metadata will be generated from the object (e.g. format, size, and other information in file headers that get there at the creation of the file), there may be a need for catalogers to verify its accuracy. There is much discussion about something called "significant properties" that are special features of the object that need to be preserved over time, and this is where it is important for an assessment to be made, and catalogers may need to play a role. An example might be a PDF file that has an embedded graph where the color of the lines determine the meaning. In this case, it is important that the color be preserved, and someone would need to look at the object to make the determination and state it in the metadata.



This is an overview of some of the standards that have been developed for technical metadata for images. We will treat NISO Z39.87/MIX in the next few slides.

XMP allows for embedding metadata in image files

Exif is used by camera companies to include metadata in file headers of digital photos. XMP metadata may also be embedded in image files.

IPTC is used by newspapers for digital photos. Might libraries be called upon to preserve any of these? That is an open question now, but we are seeing more cases of libraries becoming “trusted digital repositories” for other institutions.

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

- Includes
  - Characteristics that apply to all or most object types, e.g. size, format (elements also in PREMIS)
  - Format specific metadata for images
- Some examples of format specific metadata elements in MIX:
  - Image width
  - Color space, color profile
  - Scanner metadata
  - Digital camera settings
- Most well developed of format specific technical metadata standards

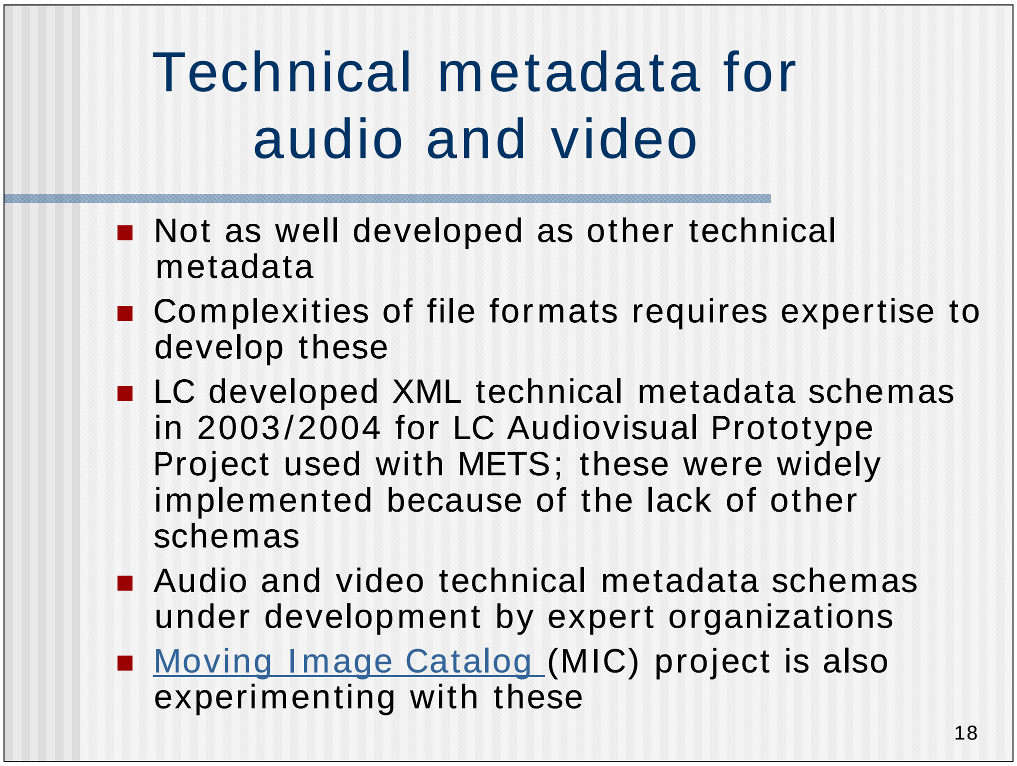
16

There is some overlap with PREMIS elements.

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



Implementers are still struggling with format specific metadata for audio and video. Although out of date, there is still much implementation of the LC prototype schemas for lack of anything else that is understandable.

The MIC Union Catalog is experimenting with technical metadata for moving image using MPEG-7. This project brings together in one place catalog records for individual moving images collected and managed by a number of organizations, enabling users to search for moving images across multiple collections. Participating institutions include LC, NLM, CNN, National Geographic Television and Film Archive, Academy of Motion Picture Arts and Sciences Academy Film Archives and a number of others.

Technical metadata for multimedia (MPEG-7)

- A multimedia content description standard, associated with the content itself
  - Intended to allow fast and efficient searching
- Formally called Multimedia Content Description Interface
  - Does not deal with the actual encoding of moving pictures and audio (as MPEG-1, MPEG-2 and MPEG-4 do)
  - intended to provide complementary functionality to the previous MPEG standards

19

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

20

We will go into structural metadata in more detail when we talk about the METS structural map in session 4.

Rights metadata

- Rights schemas with limited scope
- 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

21

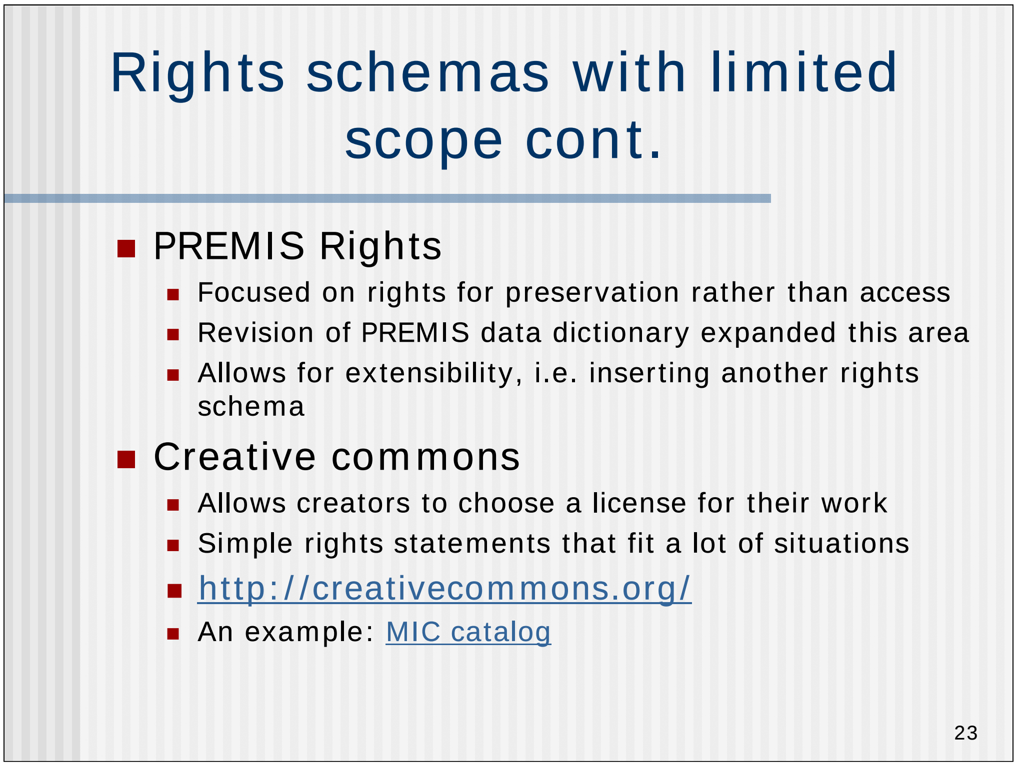
Rights is a difficult area, so there are a variety of efforts, some limited in scope, and others more expansive. Some rights are covered by laws, like copyright law, but there are also rights (also known as terms and conditions) that are governed by contracts between parties.

Rights schemas with limited scope

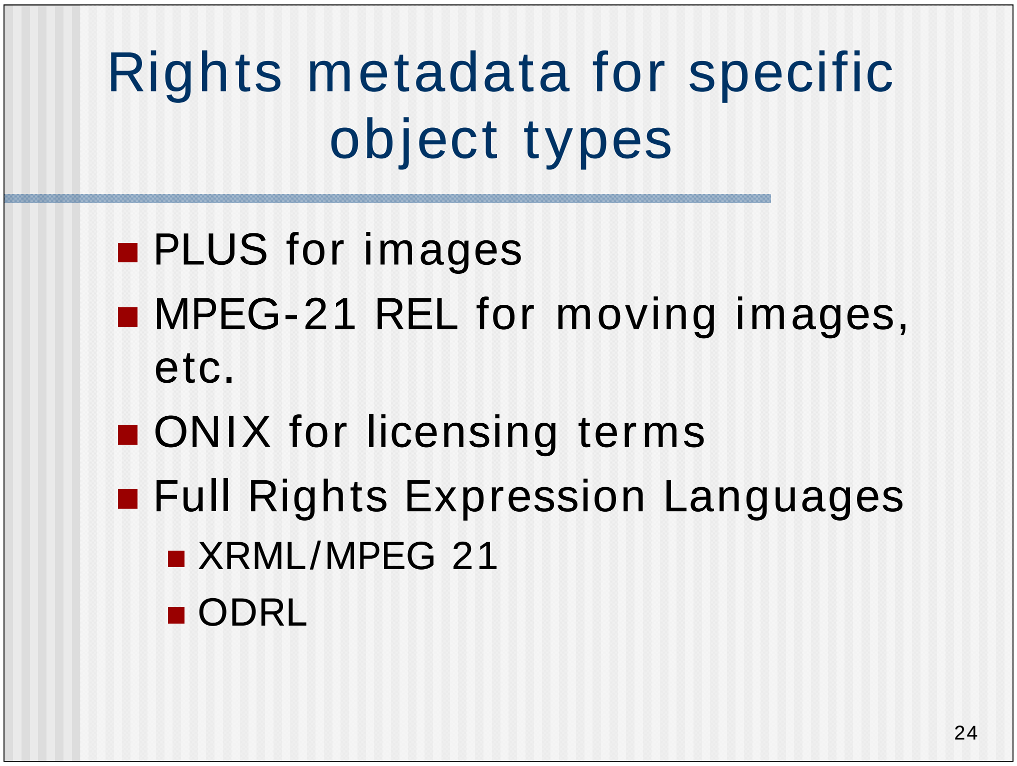
- METS Rights
  - Access rights for use with METS objects
  - Rights declarations
  - Rights holder
  - Context
- CDL copyright schema
  - Specifically copyrights, not other intellectual property rights
  - Information you need to know to assess copyright status (e.g. creators, rights holders, dates, jurisdiction)
- Note that a new field 542 has been added to MARC 21 with information about copyright to help the cataloger assess the status of the item (based on the CDL work)

22

CDL was developed by the California Digital Library for digital projects.



See bottom of page of MIC catalog: “some rights reserved” is a creative commons license.



There are also efforts to define rights metadata standards for specific object types.

The PLUS coalition is an international non-profit initiative on a mission to simplify and facilitate the communication and management of image rights. It includes a variety of members and is beginning to emerge as a standards body.

MPEG-21 is from the Motion Picture Experts Group that has developed many standards, including one which is a full rights expression language.

We talked about ONIX in terms of descriptive metadata, but it also is concerned with rights issues and has developed a list of licensing terms.

ODRL is a competitor of XRML/MPEG 21 (Open Digital Rights Language). Its major difference is that it is an open standard, while XRML/MPEG 21 is proprietary and requires payment for use.

Exercise

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

25

We will take the same objects used for the descriptive exercise and fill in some PREMIS elements with the information provided in the session 3 exercise handout.

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

Example 1 (sheet music):

This is a digitized version of 3 pages of sheet music. The first page is the cover. There are 3 files, one for each page. The files use the JPEG format.

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):
FileID: 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):
FileID: 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):

This item was digitized in the TIFF format, which is a high quality master format used as a preservation copy. A derivative was made in the JPEG format, which is considered a "service" copy (used for retrieval). There are 2 files: the TIFF master and the JPEG derivative.

52nd Street, New York, N.Y., ca. 1948
http://lcweb2.loc.gov/natlib/ihas/warehouse/gottlieb/02771/ver01/0001.tif

File 1 (master TIFF)
FileID: masterd1e30196
Full path:
http://lcweb2.loc.gov/natlib/ihas/warehouse/gottlieb/02771/ver01/0001.tif
File format: image/tiff
Size: 60158210

Software used to access: Firefox version 5.0
File created: Feb. 15, 2003
Application used for creation: Image Alchemy PS v1.11
Date created: 20030215

File 2 (derivative JPEG)
FileID: serviced1e30196
Full path:
http://lcweb2.loc.gov/natlib/ihas/service/gottlieb/02771/ver01/0001v.jpg
File format: image/jpeg
Size: 167367
Software used to access: Firefox version 5.0
File created: Feb. 15, 2003
Application used for creation: Image Alchemy PS v1.11
Use metadata template provided.

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.

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

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

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

4. Metadata syntaxes and containers

Metadata Standards and Applications Workshop

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

[Point out the handout that defines some of the XML terms.]

It may not be necessary to understand these syntaxes in detail unless you are involved in using them for processing or encoding, since people don’t normally need to look at the raw data. An understanding is helpful so that you know how information can be transformed from the raw syntax to a user friendly display using widely available tools. As with learning the syntax of MARC records, for instance, it is also useful for problem solving when information doesn’t display as expected.

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

SGML has not translated well to the web and has generally been succeeded by XML.

Use HTML for simple documents and web presentation.

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

HTML does not require ending tags. XHTML makes HTML look more like XML. It is important to know that XHTML is a subset of XML to know what it works for and what it does not work well for.

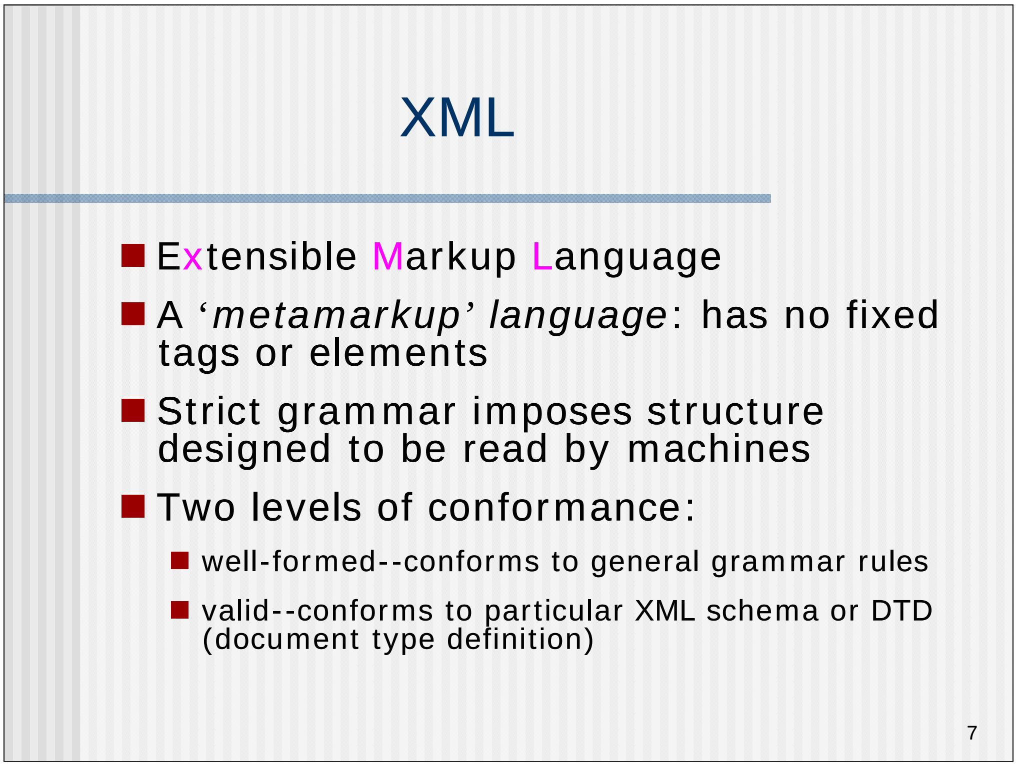
An XHTML Example

```
<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/" />
```

This example was created using the DC-Dot metadata generator. Note that it has a link to 2 namespaces: DC elements, DC terms.

Note also the variable usefulness of the subject keywords.

XHTML is commonly used for digital objects on the web



XML on its own is a flexible syntax that doesn't have a lot of specified elements. You use whatever element sets you want to use and define them in an XML schema. The structured grammar is for machines, not to be read by humans, although some of us have adapted to being able to read it.

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

More on XML. Tags can be defined in any language and are most often codified in a schema. There is no requirement to use English tags, for instance.

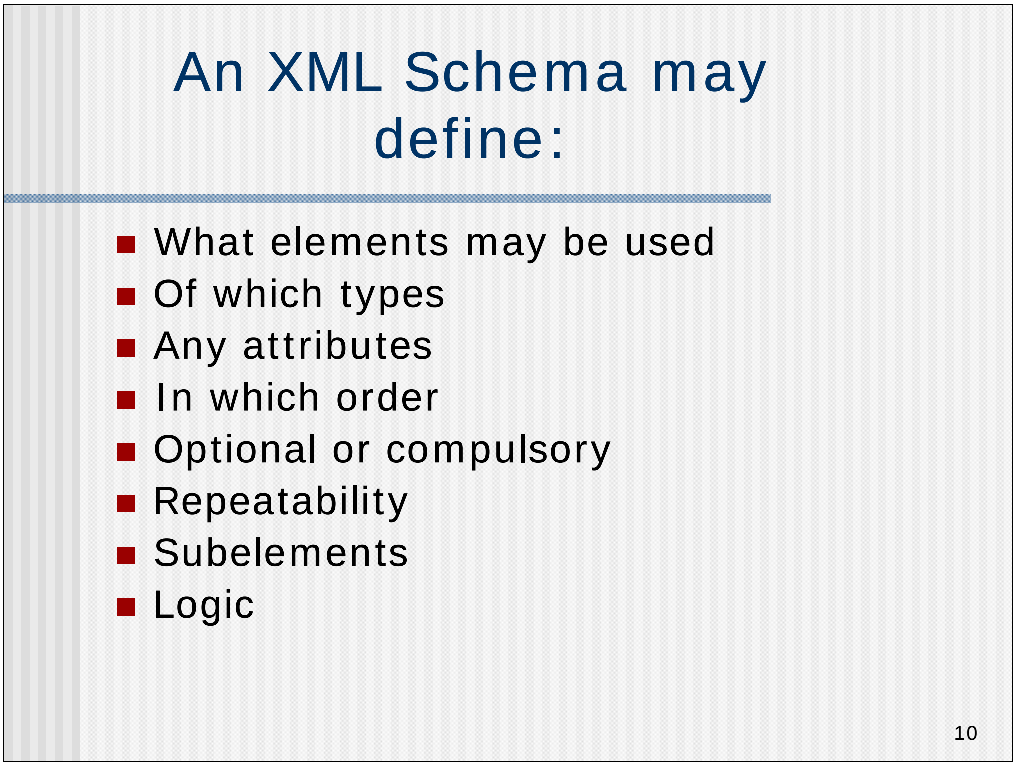
XML is the *lingua franca* of the Web

- 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

There are lots of freely available tools for processing XML documents and using them for different purposes. XSLT is for XML "stylesheets", which tell a computer how to process and/or display an XML document. It is used for many purposes, but particularly to take raw XML and display it in a user friendly form, perhaps converting from XML to HTML for display on the Web. Xpath is for linking and Xquery is for searching.

It is also becoming a preferred standard for encoding metadata (although it is not the only one).



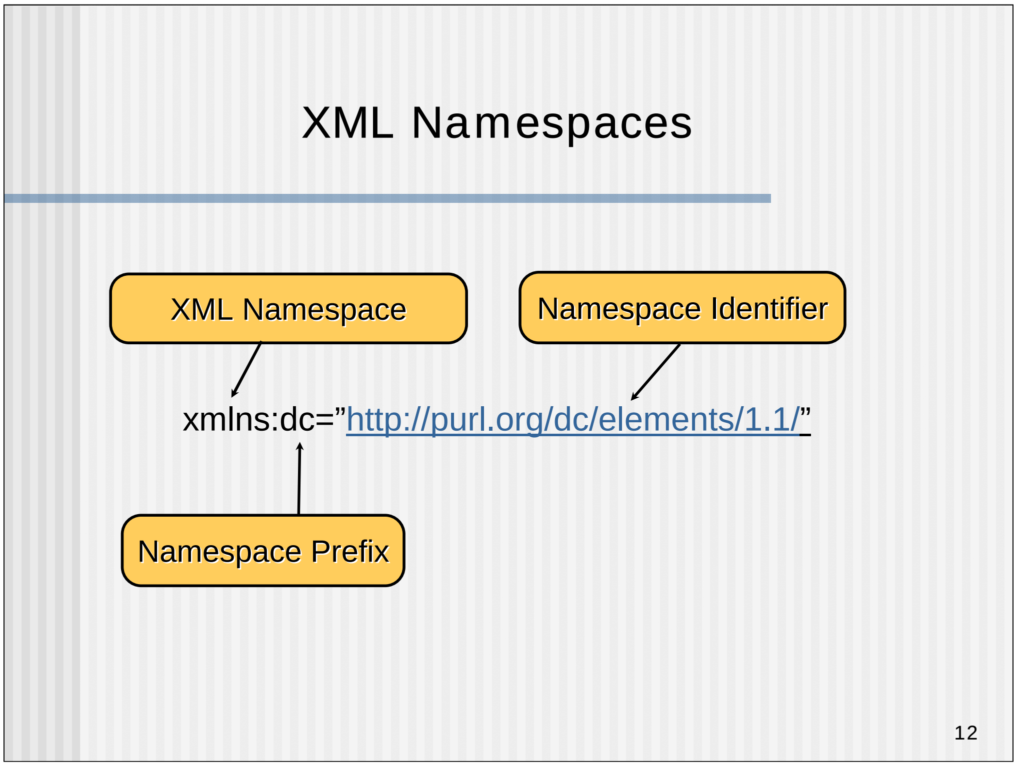
Since XML is a common format for metadata (according to a defined XML schema), it is helpful to know what XML schema defines.

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

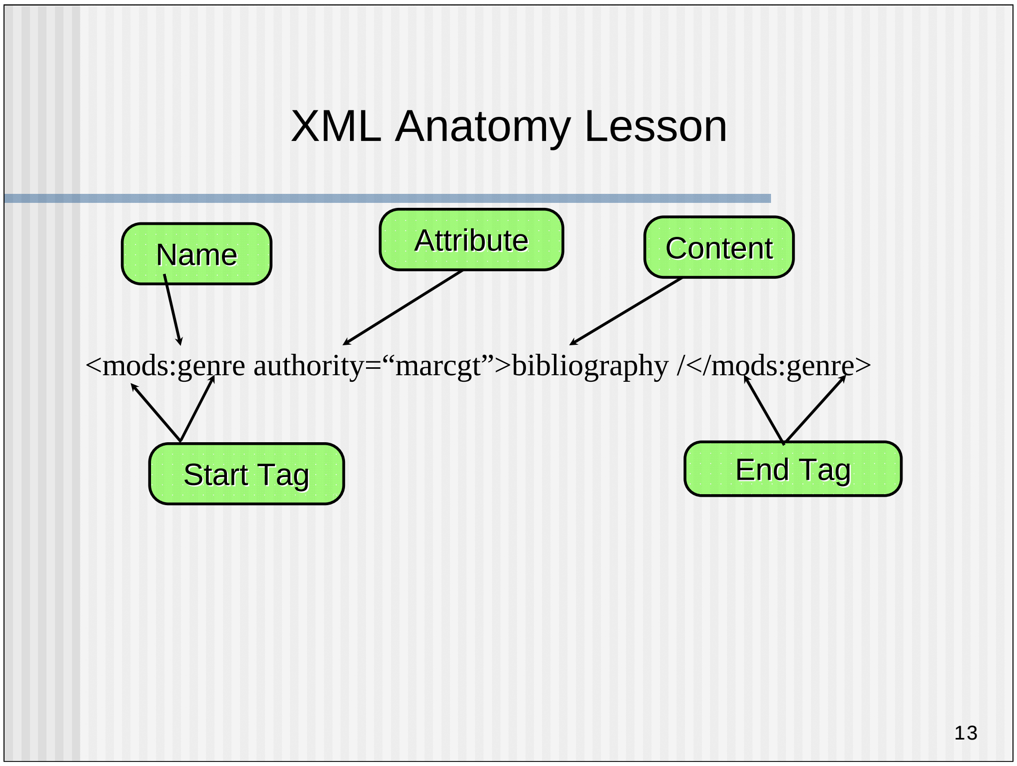
11

The URI of a namespace might not be resolvable– there might not be something at the other end, but the important thing is that it is there to resolve the ambiguity.

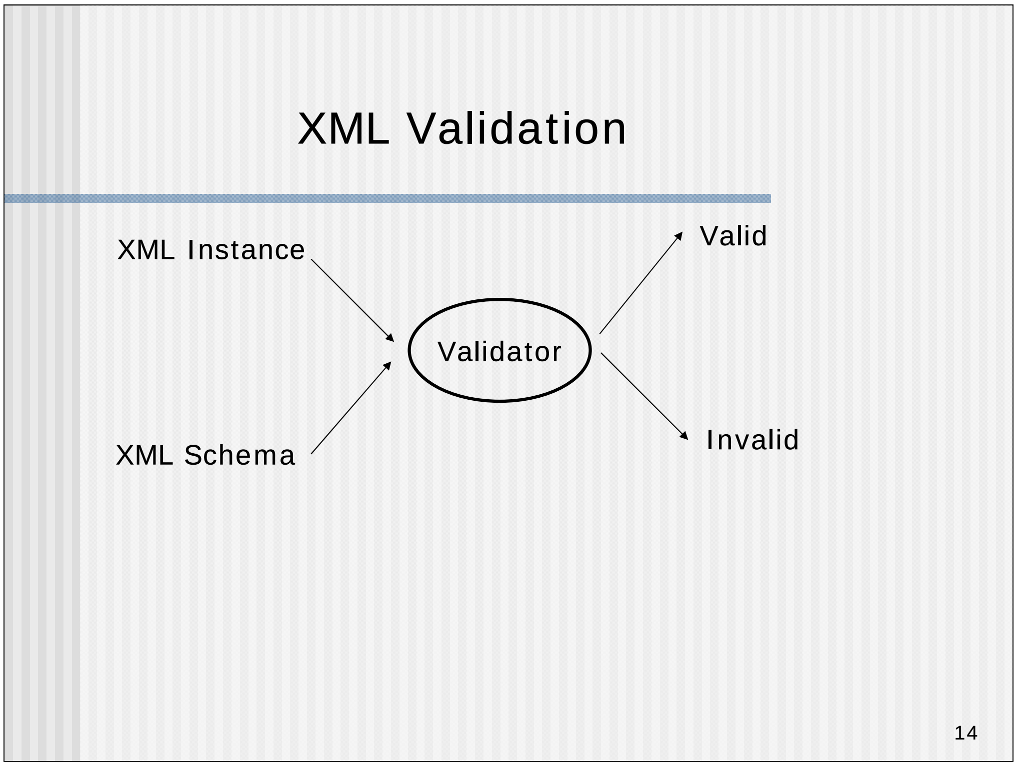


"xmlns" is defined by the XML standard to encode the XML namespace where the schema is found.

The namespace prefix is what is used to distinguish where a given XML element comes from.



The pieces that come together to build an XML statement (nested within another element). For each start tag there must be an end tag. The content is the value between the tags. Attributes may be used for further information about the given tag.



Validators can check where the XML record (here seen as "XML instance") is valid according to the schema. These are part of any XML editing software.

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>
```

This XML schema example is from the PREMIS environment section.

minOccurs specifies whether it is mandatory or optional (minOccurs=“0” is optional; minOccurs=“1” is mandatory

maxOccurs specifies repeatability (maxOccurs=“1” is not repeatable; maxOccurs=“unbounded” means it can be repeated an unlimited number of times)

complexType defines additional structure of the element.

xs:sequence specifies order

Type specifies a data type (e.g. string is text; integer is number; xs:date is a special structured form of date)

Will the following XML instance validate?

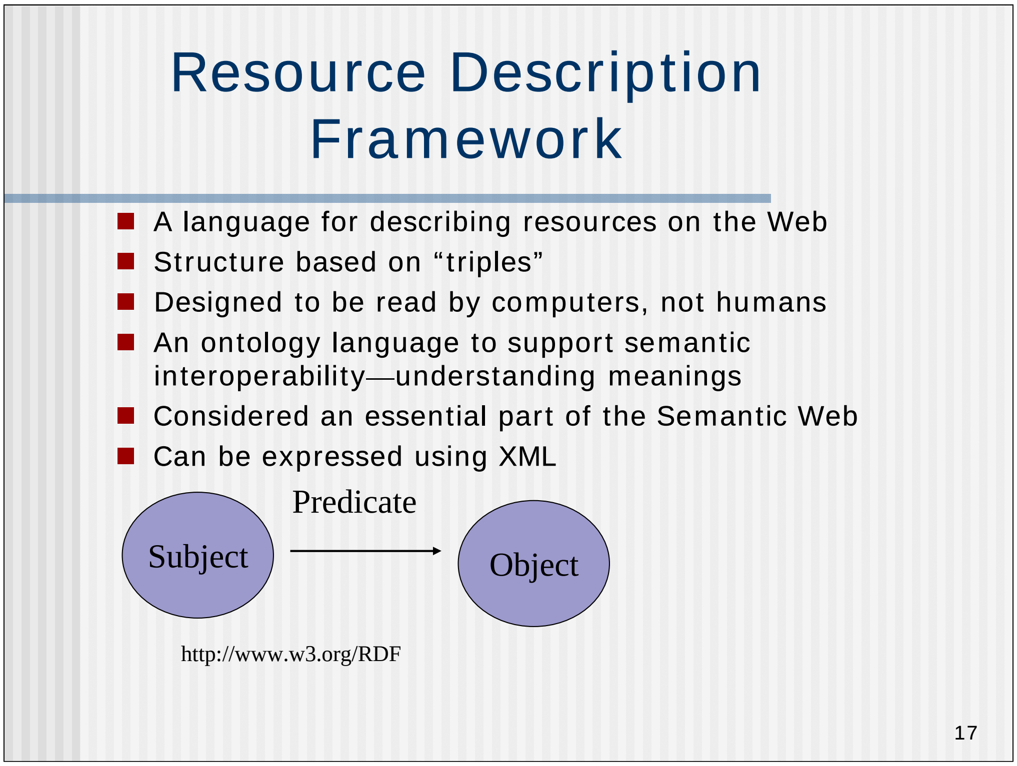
```
<software>
   <swName>Windows</swName>
   <swVersion>2000</swVersion>
   <swType>Operating System</swType>
</software>
```

How about this?

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

16

The second one won’t validate if the schema specifies that the element is in a container and the container element is not included. It will only validate if <software> start and end tag is there.



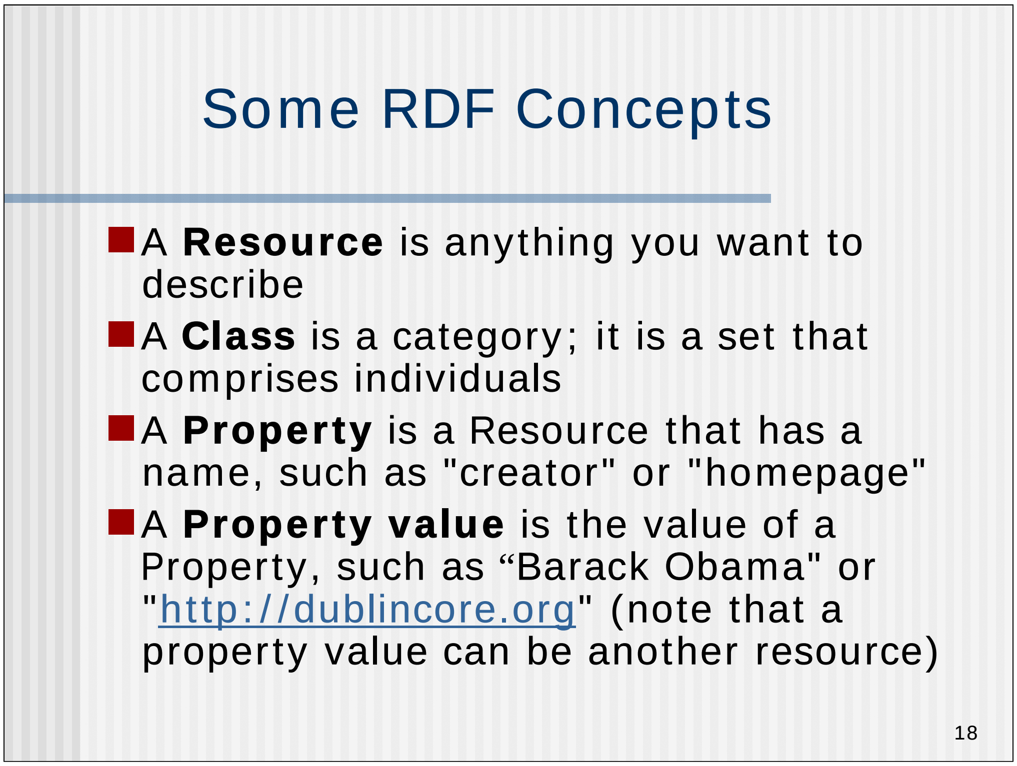
RDF was created in 1999 as a standard that sat on top of XML; has evolved away from encoding metadata to describing the relationships between concepts (names (people, corporate bodies, place), topics, objects

Subject = Topic

Predicate = Relationship (hierarchical, equivalent, associative)

There are a few methods to encode RDF, including in XML.

Permits infinite number of combinations which can provide a framework for the development taxonomies (thesauri, controlled vocabularies, faceted classifications)



Simple RDF concepts.

Something that is an element-property can also be considered a conceptual resource, very much like a controlled vocabulary term.

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

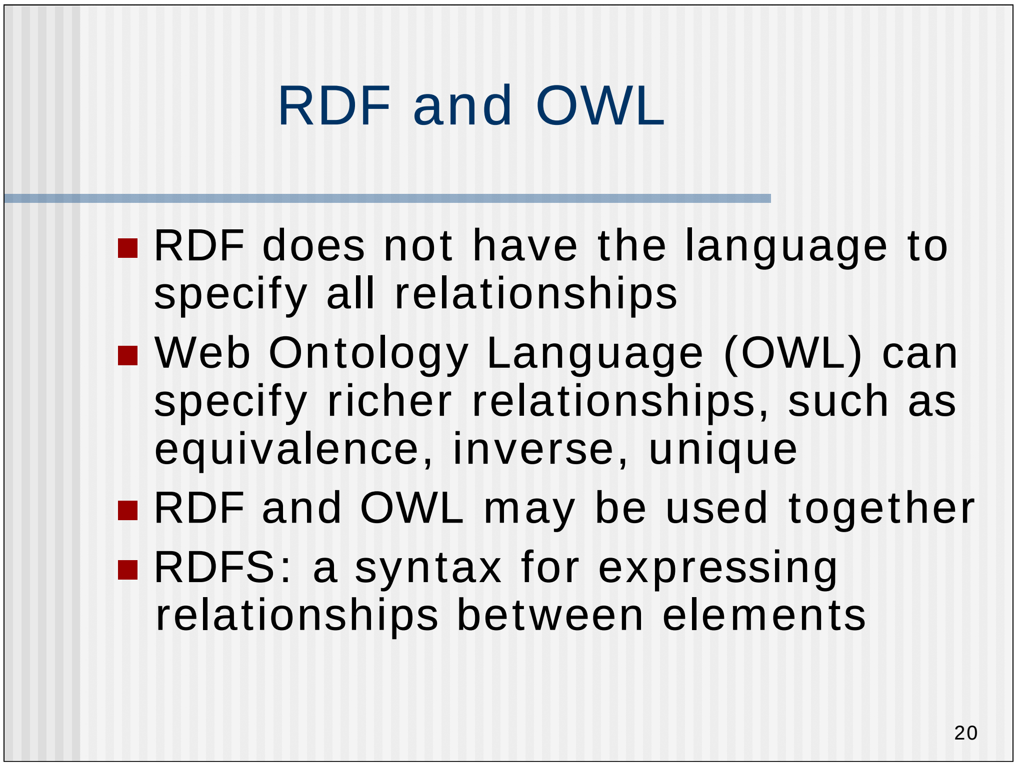
Resource = subject

Property = predicate

Property value = object

… of an RDF statement.

These three pieces are called “triples.”



OWL is still in development and there aren't a lot of tools yet, although it has potential. RDF and OWL are all about defining relationships between things. RDF syntax is limited in the relationships it can express. For instance there is nothing like what we have in our library authority files that mean a "see" reference, that is, here is an alternate name, but don't use this one, used the preferred one.

RDFS is another syntax for RDF which is a schema itself (not the same as RDF/XML).

An RDF/XML Example

```
<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:publi
     <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

"Native" RDF does not need XML, but in a world that is much more comfortable with XML "packages," it's a useful way to move data around.

In this RDF/XML version of a DC record, the resource URI is expressed as rdf:about rather than dc:identifier. In both cases, the other statements (expressed here as attribute/value pairs) can be transformed into triples by associating the resource URI with each attribute/value pair.

Under <subject>, we have a list. You can make an ordered list as well, called "seq" in RDF.

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

Now we will turn to the use of container formats, which themselves require a syntax.

OAIS=Open Archival Information System, a widely used standard for understanding digital repositories and functions

METS is more prevalent in digital library applications, although MPEG 21 Document Item Declaration (DID) is an alternative.

Metadata Encoding & Transmission Standard (METS)

- 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

METS was developed in response to a need to encode metadata and information about digital files in information packages.

METS Usage

- 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

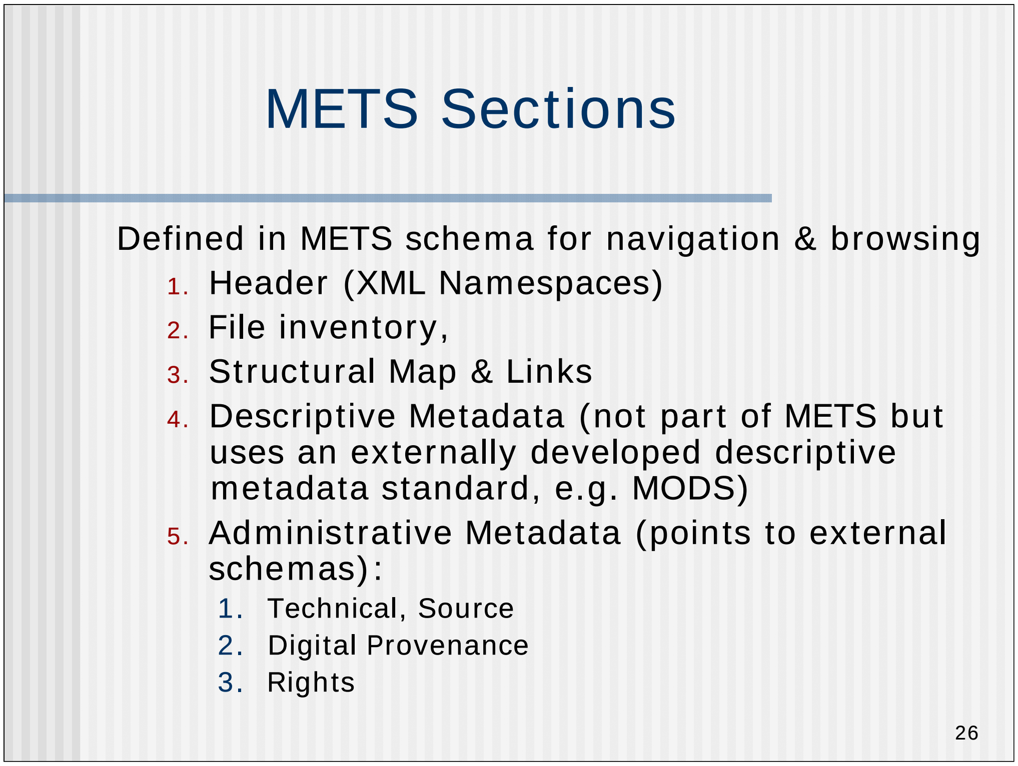
OAIS defines 3 types of information package: Submission Information Package, Archival Information Package (storage), and Dissemination Information Package (transfer). METS can function as any of these

Characteristics of METS

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

25

METS Editorial Board endorsed extension schemas for bibliographic data (MARCXML, MODS, DC), for technical metadata for still images (MIX), and technical metadata for text.

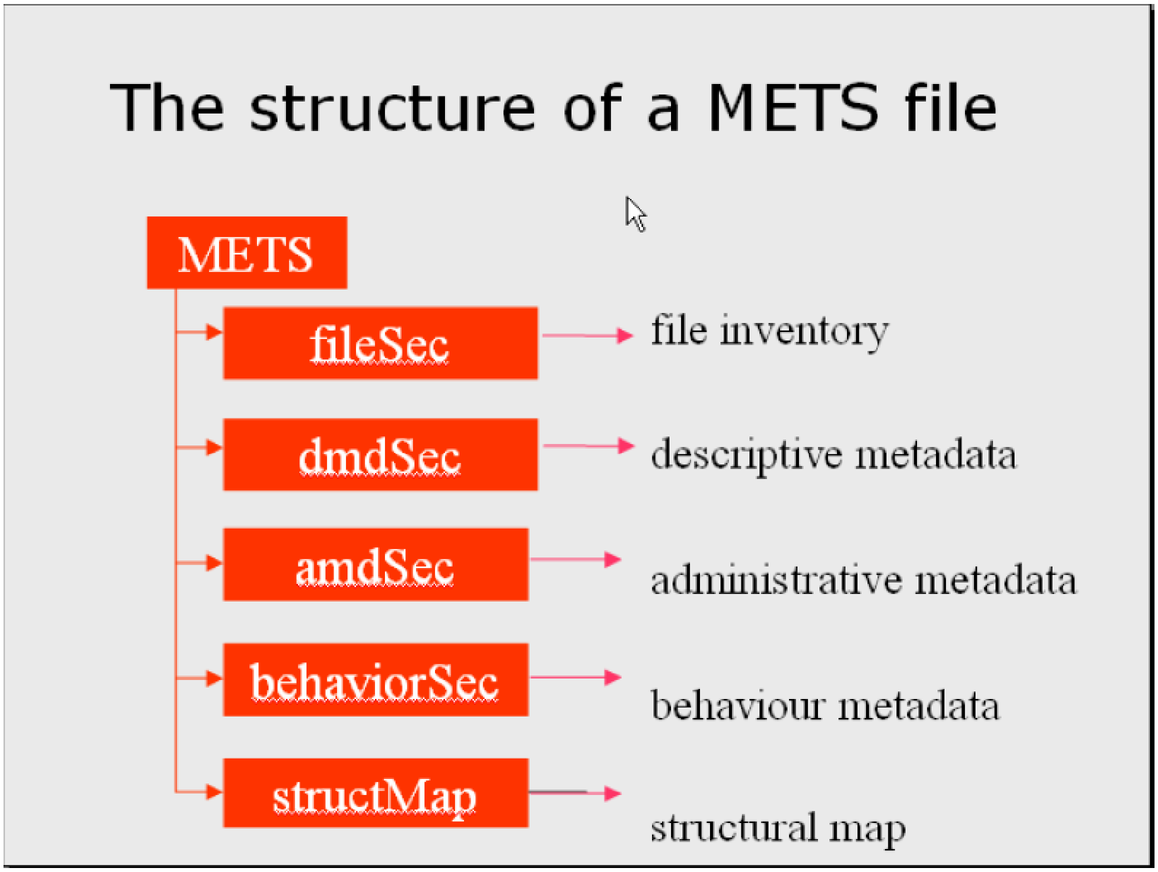


Descriptive metadata describes the package itself. The METS document can be at any level of granularity in terms of the intellectual object it represents.

The METS schema only defines the first 3 sections.

File section: includes a list of all the files in the METS package

Structure map: how the structure of the object is defined; links the digital object to content files and the metadata.



A graphical view of the structure of METS. Behavior section was added later; there are not a lot of uses of it yet.

Linking in METS Documents
(XML ID/IDREF links)

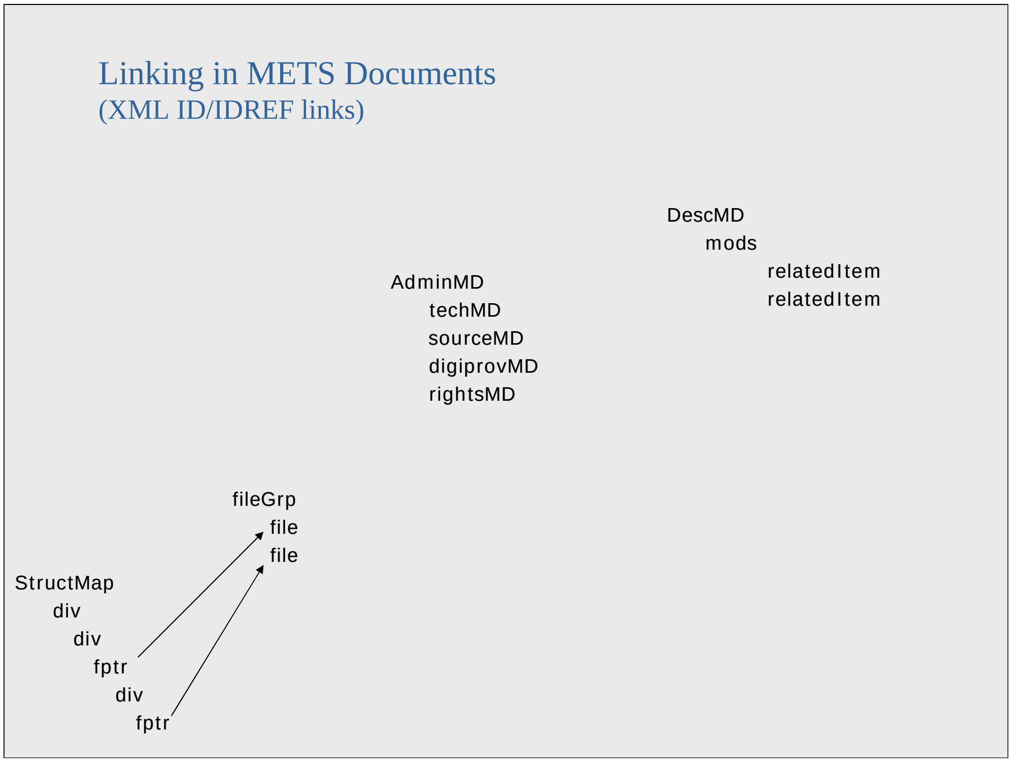
DescMD
mods
relatedItem
relatedItem

AdminMD
techMD
sourceMD
digiprovMD
rightsMD

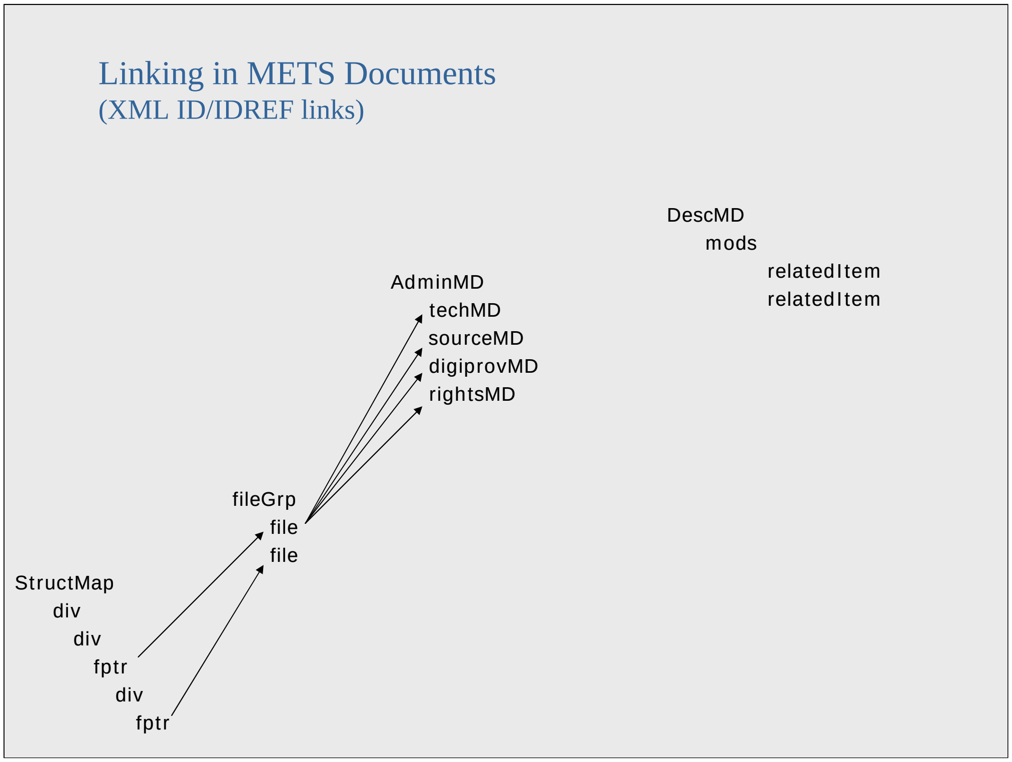
fileGrp
file
file

StructMap
div
div
fptr
div
fptr

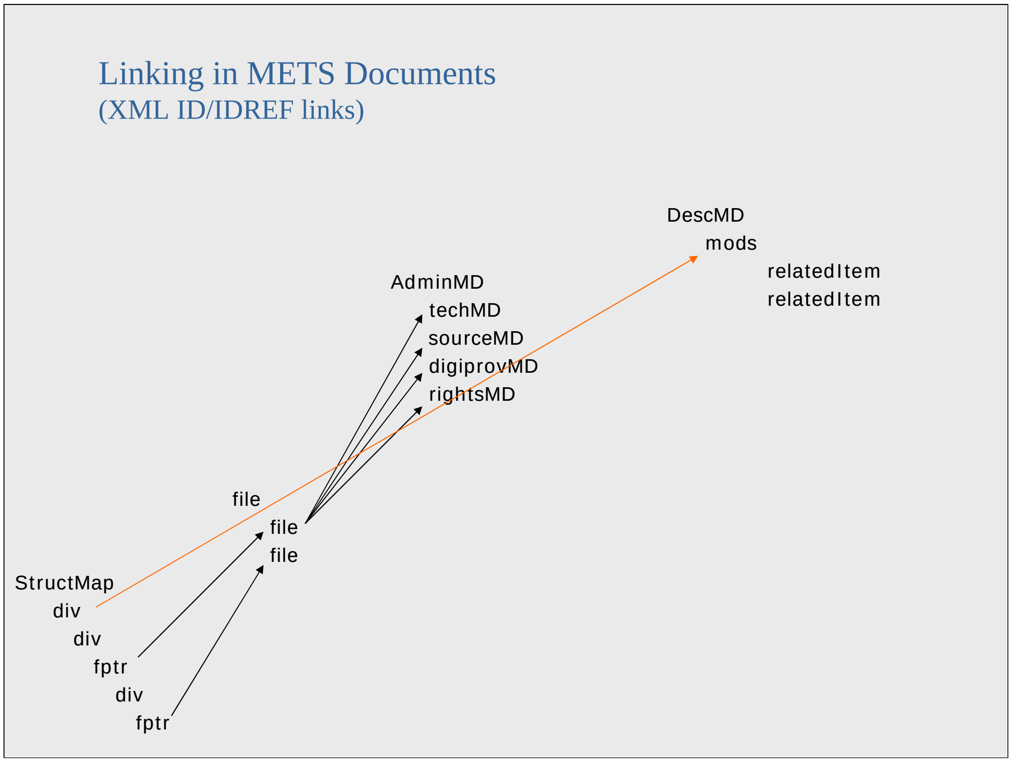
The structMap (structural map) is the heart of the METS document and specifies its structure by using nested “div” elements. This shows how objects consisting of multiple files are put together. To understand METS and how it uses metadata you need to understand how linking is done. The <div>s show hierarchy. “fptr” is called the file pointer.



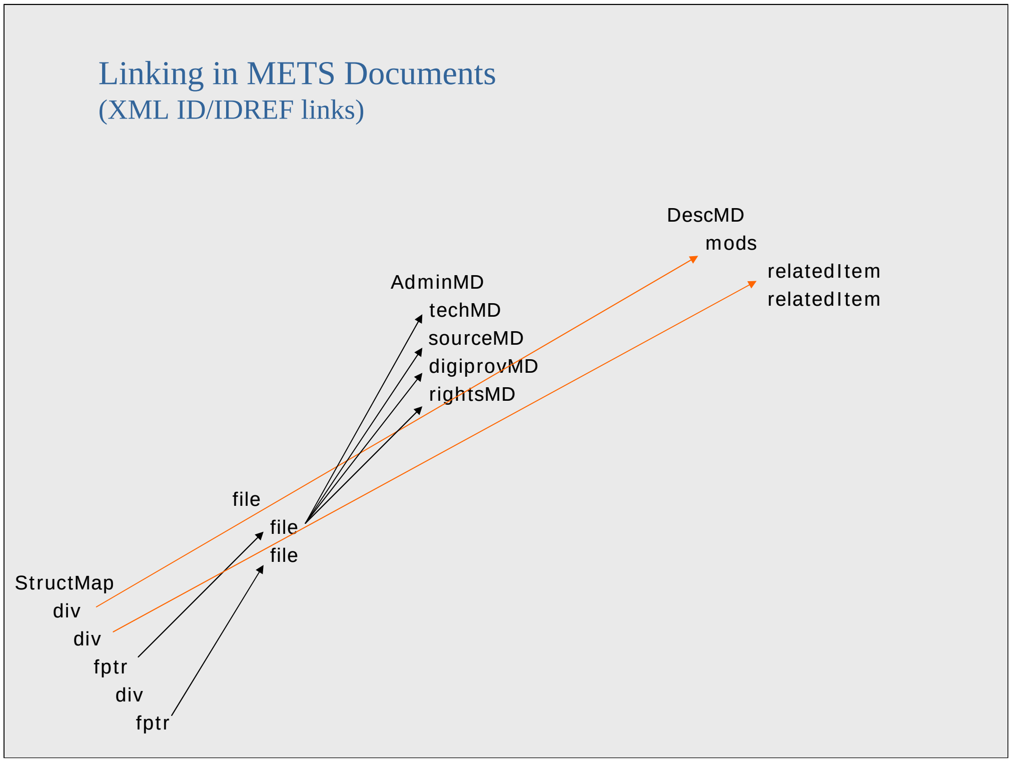
From the various hierarchical parts of the structural map you point to the relevant files in the file group.



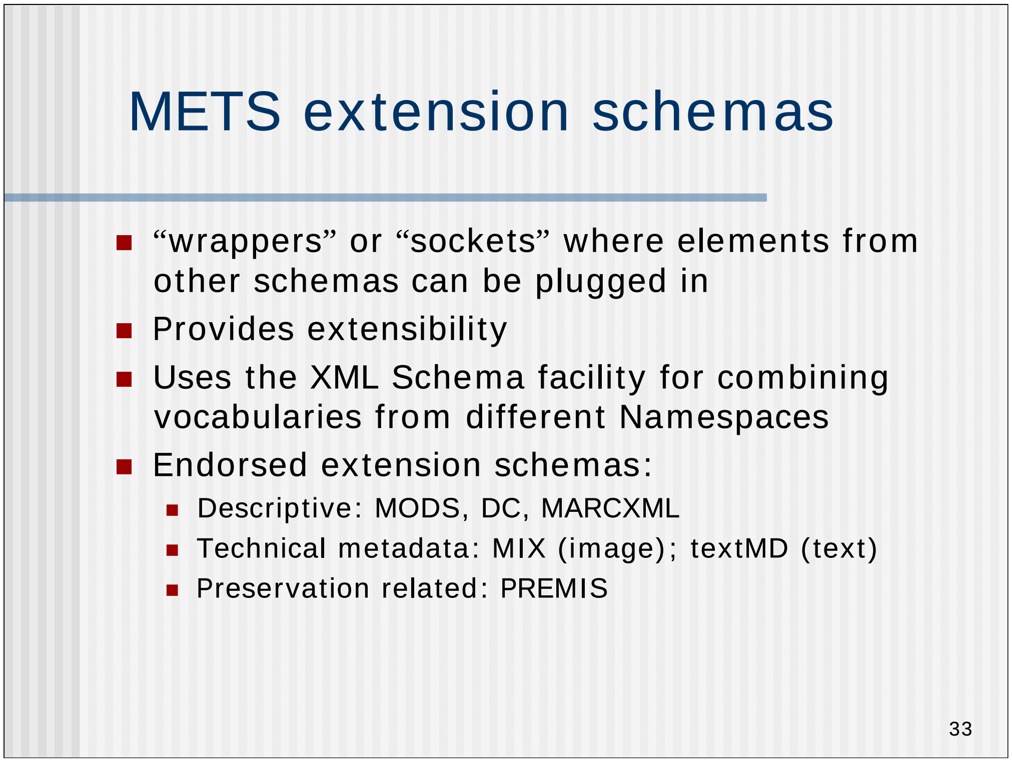
From the files you can point to the administrative metadata sections that apply to the particular file. Each file may be linked to multiple metadata sections.



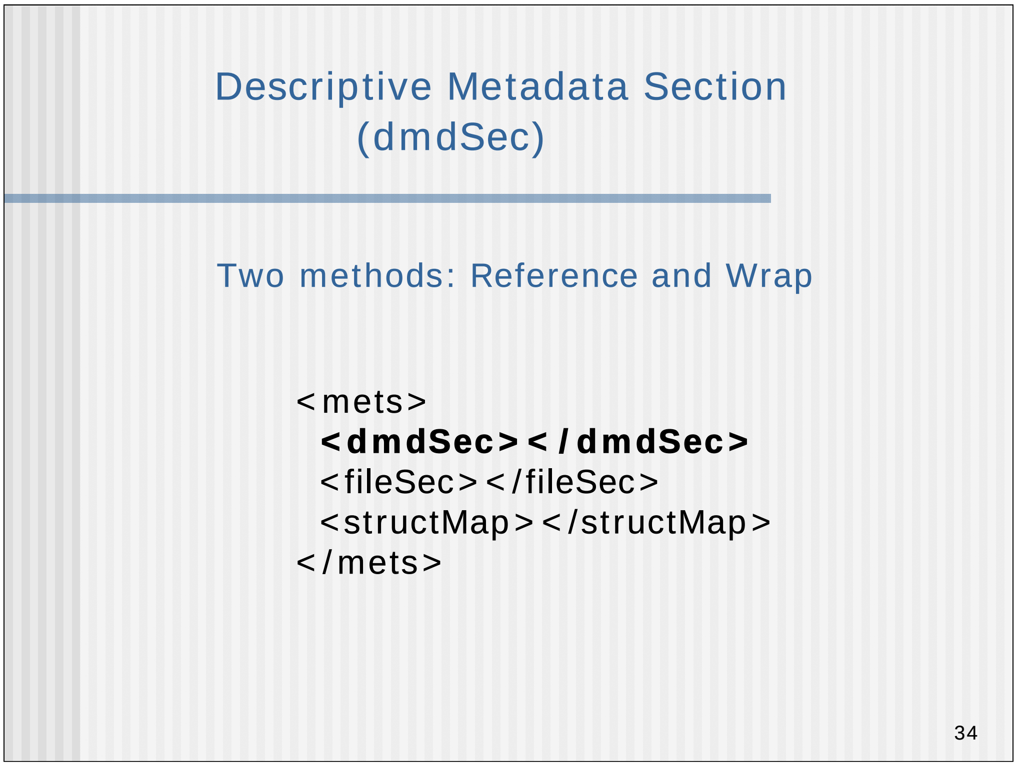
You can link to descriptive metadata in the descriptive metadata section for the object as a whole or for parts of the object.



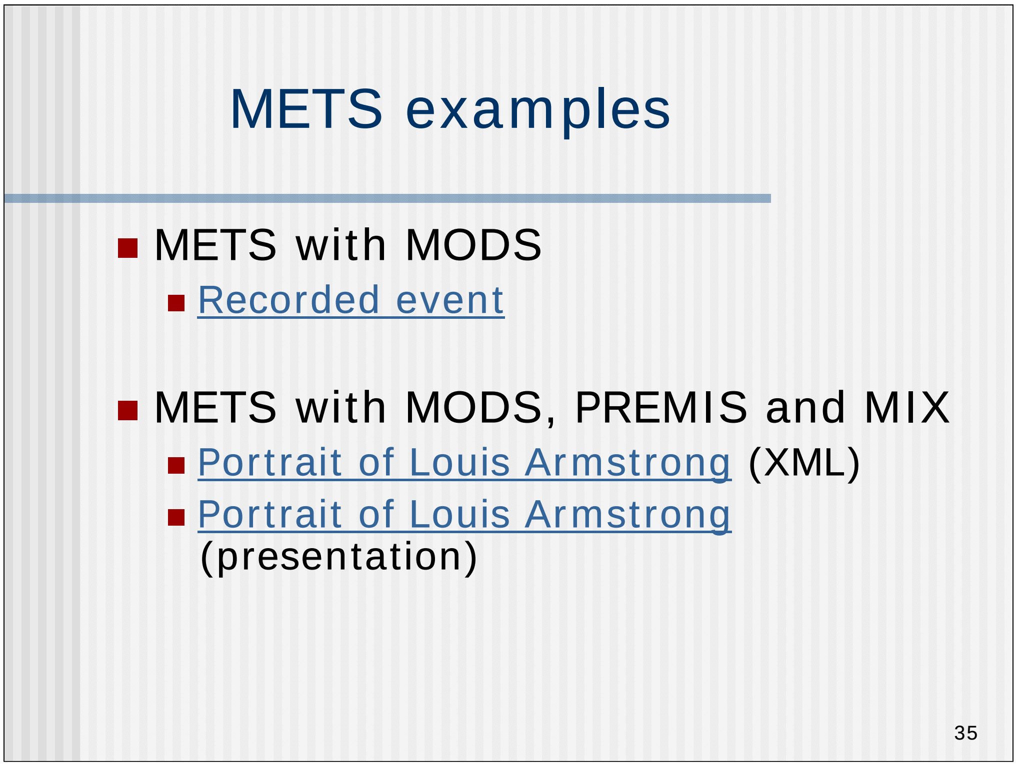
This is where METS and MODS work well together because both are hierarchical. For instance, if you have a CD with 3 tracks, you can specify metadata (both descriptive and administrative) for each track by using MODS relatedItem type="constituent". We will see an example of this in the next session.



The METS Editorial Board has endorsed a number of extension schemas, although any XML format may be used.



You can insert the metadata in the METS document itself (as XML) or you can reference it, for instance as a record in a library catalog.



Link to Recorded event. A METS document is behind the entire display that you see. Click on the very bottom of the page where it says “METS Object Description”. Point out the structural map with the various levels of the hierarchy. The image on the right is displayed using an XSLT stylesheet which specifies display of one of the files in the file section of the METS document (the thumbnail version). The MODS record is displayed with field names as specified by the stylesheet. Additional files may be accessed or displayed from the other linked files in the file section (e.g. the concert program under “See also”).

.

The Louis Armstrong examples shows PREMIS metadata within METS in addition to descriptive metadata.

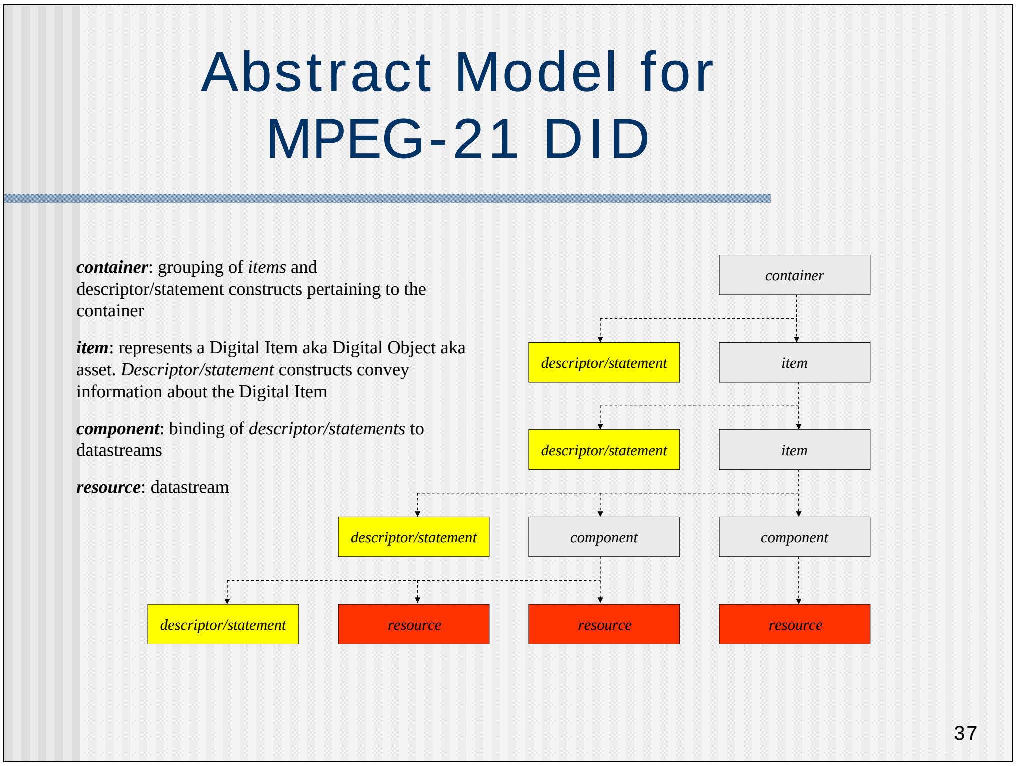
Click on Portrait of Louis Armstrong (XML) to show where PREMIS is included in the METS document (under amdSec/techMD and amdSec/digiProv MD). PREMIS object information (technical metadata) is under techMD and information about events (ingest and validation) is in digiProv. Linking between files to the metadata is done with the ID attribute on the files to the appropriate metadata section.

If you want show the display under Portrait of Louis Armstrong (presentation); this is derived from the METS document.

MPEG-21 Digital Item Declaration (DID)

- ISO/IEC 21000-2: Digital Item Declaration
- An 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”)
- MPEG DID is an ISO standard and has industry support, but is often implemented in a proprietary way and standards development is closed; METS is open source and developed by open discussion, mainly cultural heritage community

36

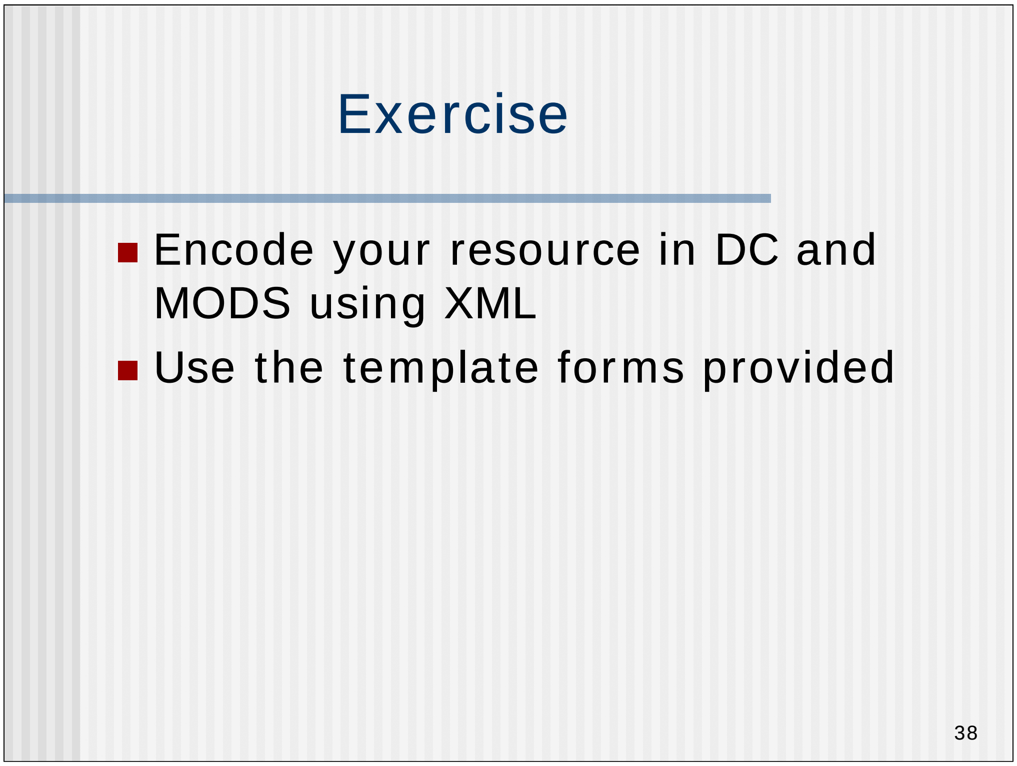


METS and MPEG DID are similar types of container formats in that both are expressed in XML, both represent the structure of digital objects, and both include metadata.

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.

It would be possible to transform a METS container to a MPEG DID and vice versa; development of stylesheets will enable transformations



Use the same object you used for the descriptive and technical metadata and the subset of elements provided in the exercise handout. You may need to add additional fields to repeat the tags as needed.

Exercise for Session 4: XML Syntax
DC Template

```
<?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/">

   <dc:title xml:lang="    ">


   </dc:title>

   <dc:creator>


   </dc:creator>

   <dc:type>


   </dc:type>

   <dc:publisher>


   </dc:publisher>

   <dc:date>


   </dc:date>

   <dc:format>


   </dc:format>

   <dc:identifier>

   </dc:identifier>

  <dc:subject xsitype="dcterms:LCSH">


  </dc:subject>

 <dc:relation>
```

```
  </dc:relation>
</metadata>
```

Exercise for Session 4: XML syntax
MODS Template

```
<?xml version="1.0" encoding="UTF-8"?>
<mods:mods version="3.2" ID="MODS1"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/mods.xsd">

  <mods:titleInfo>
     <mods:title>


     </mods:title>
     <mods:subTitle>


      </mods:subTitle>
  </mods:titleInfo>

  <mods:name type=”personal”>
    <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:dateIssued>
```

```
     </mods:dateIssued>

     <mods:dateCreated>


     </mods:dateCreated>
   </mods:originInfo>

   <mods:physicalDescription>
       <mods:form authority=" ">sheet music</mods:form>

       <mods:extent>


       </mods:extent>
    </mods:physicalDescription>

  <mods:accessCondition>


  </mods:accessCondition>

   <mods:subject authority="lcsh">
       <mods:topic>


      l</mods:topic>
       <mods:genre>


       </mods:genre>
   </mods:subject>

  <mods:subject authority=" ">
      <mods:name>
         <mods:namePart>


         </mods:namePart>
      </mods:name>
    </mods:subject>

  <mods:physicalLocation>
     <mods:url>


     </mods:url>
  </mods:physicalLocation>

  <mods:relatedItem type="host">
       <mods:titleInfo>
          <mods:title>
```

```
        </mods:title>
      </mods:titleInfo>
    </mods:relatedItem>

</mods:mods>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<mods:mods version="3.2" ID="MODS1"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/mods.xsd">

  <mods:titleInfo>
     <mods:title>


     </mods:title>
     <mods:subTitle>


      </mods:subTitle>

  </mods:titleInfo>

  <mods:name type="personal">
    <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:dateIssued>


    </mods:dateIssued>
```

```
      <mods:dateCreated>


      </mods:dateCreated>
    </mods:originInfo>

    <mods:physicalDescription>
        <mods:form authority=" ">sheet music</mods:form>

        <mods:extent>


        </mods:extent>
     </mods:physicalDescription>

  <mods:accessCondition>


  </mods:accessCondition>

    <mods:subject authority="lcsh">
        <mods:topic>


       l</mods:topic>
        <mods:genre>


        </mods:genre>
    </mods:subject>

  <mods:subject authority=" ">
       <mods:name>
          <mods:namePart>


          </mods:namePart>
       </mods:name>
     </mods:subject>

  <mods:physicalLocation>
      <mods:url>


      </mods:url>
  </mods:physicalLocation>


  <mods:relatedItem type="host">
        <mods:titleInfo>
            <mods:title>
```

```
          </mods:title>
        </mods:titleInfo>
      </mods:relatedItem>

    </mods:mods>
```

```
<?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/">

   <dc:title xml:lang="    ">


   </dc:title>

   <dc:creator>


   </dc:creator>

   <dc:type>


   </dc:type>

   <dc:publisher>


   </dc:publisher>

   <dc:date>


   </dc:date>

   <dc:format>


   </dc:format>

   <dc:identifier>


   </dc:identifier>

  <dc:subject xsitype="dcterms:LCSH">


  </dc:subject>

 <dc:relation>


 </dc:relation>
```

```
</metadata>
```

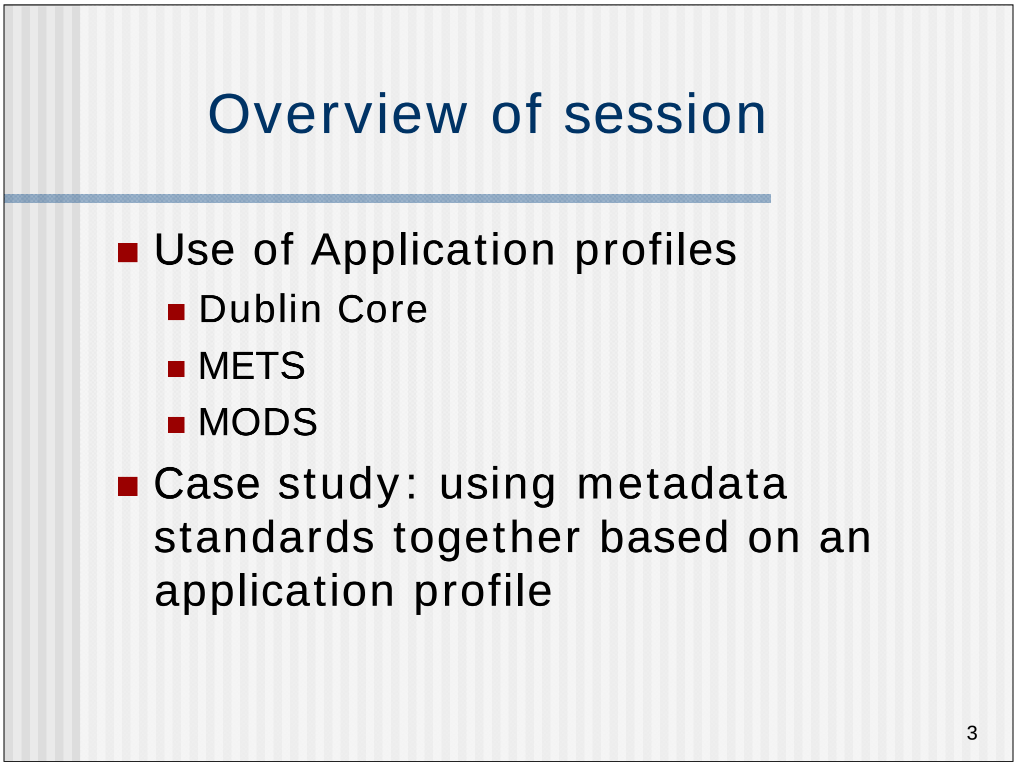
5. Applying metadata standards: Application profiles

Metadata Standards and Applications Workshop

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 may be used together in digital library applications

2


Overview of session
Use of Application profiles
Dublin Core
METS
MODS
Case study: using metadata standards together based on an application profile
3

Why Application profiles?

- Describes the set of metadata elements, policies, and guidelines defined for a particular application, implementation, or object type
  - Declares 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”

4

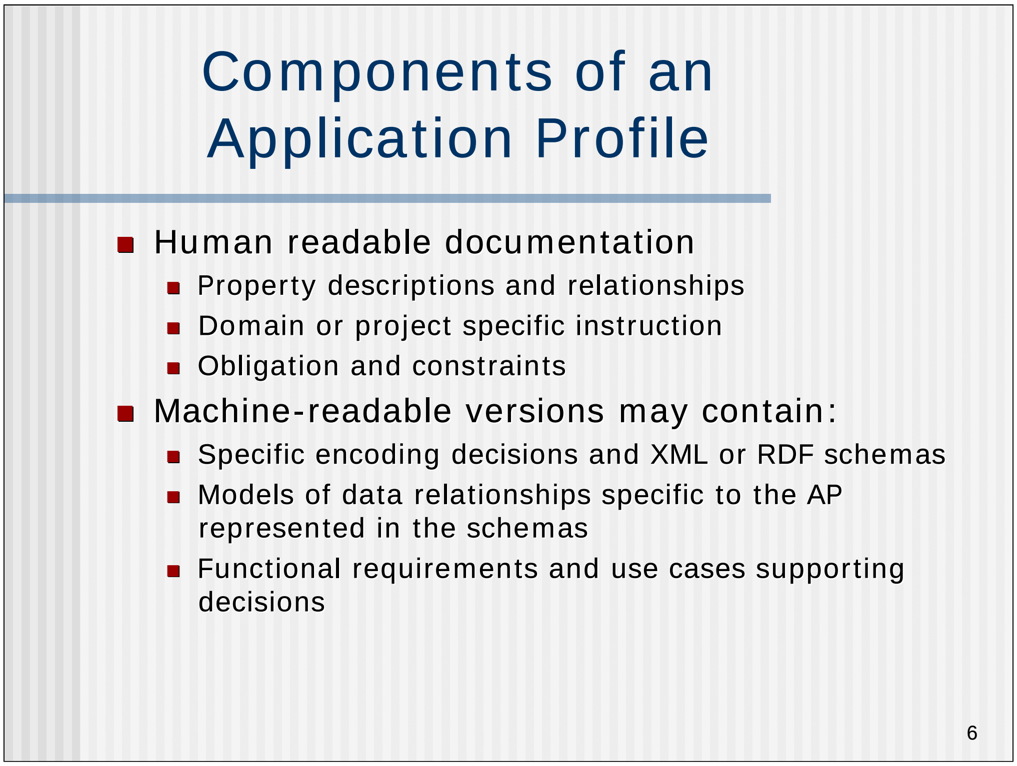
A number of different metadata efforts have begun to develop profiles for particular applications. This gives some guidance and rules for imposing some constraints on metadata standards that by nature are flexible.

Function of Application Profiles

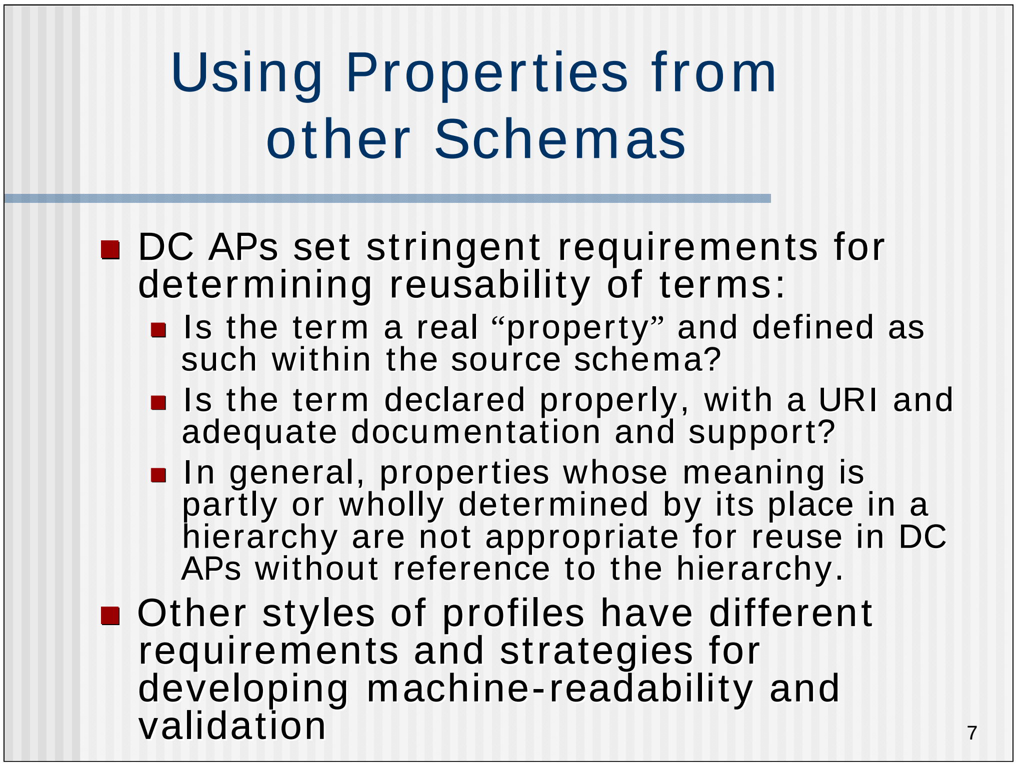
- Many metadata standards are sufficiently flexible that they need a mechanism to impose some constraints
  - Profiles allow expression of the decisions made for a project in machine-readable form (XML or RDF)
  - Profiles allow for enforcing those decisions
  - This facilitates interoperability and common practices
- Refining
  - A narrower interpretation of a standard to suit your project
- Combining
  - Mixing elements from various different standards

5

Application Profiles document the hard choices implementers make when adopting standards – for example, they reflect whether a data field is deemed mandatory or optional or simply unused for a particular project. In all of this tailoring, the application profile always remains within the limits set by the larger standard. It can narrow the scope of a standard, but it always has to still conform with the specification it originates from.

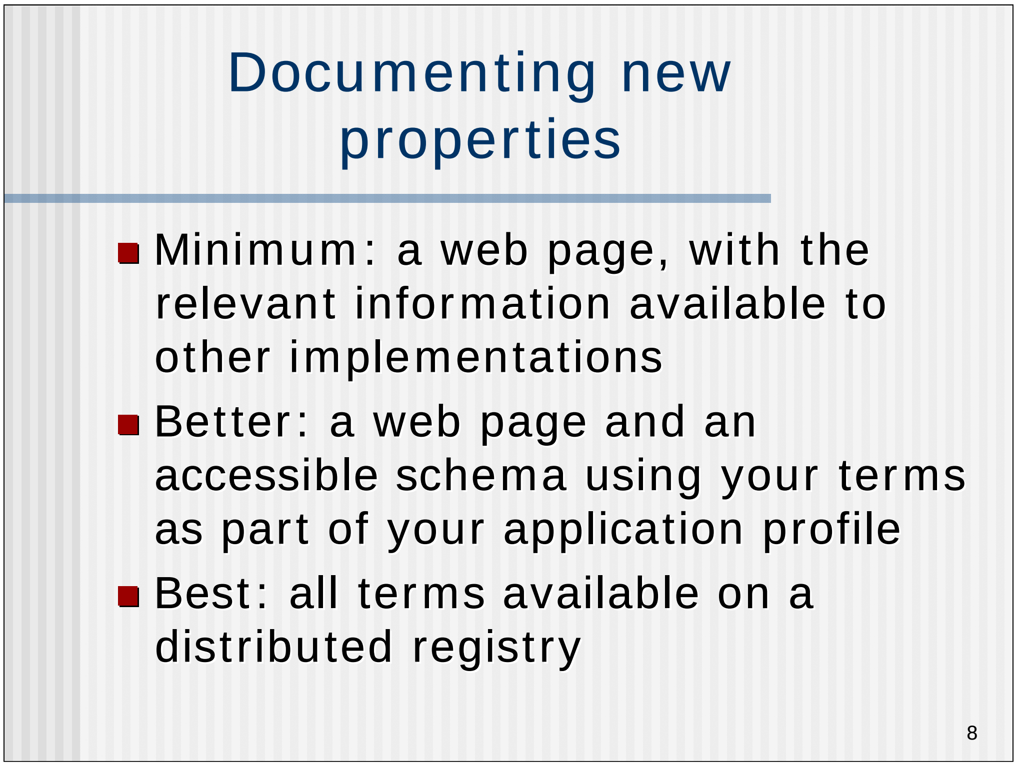


An AP makes it possible for people to share community consensus and project decision making. Machine-readability doesn't guarantee that data will be interoperable, but Application Profiles make it more likely that machines will be able to cope.



Different metadata element sets differ in terms of hierarchy in particular. For example, Dublin Core has a flat structure. MODS, though, involves a hierarchy, which is not easily expressed in a Dublin Core Application profile problematic.

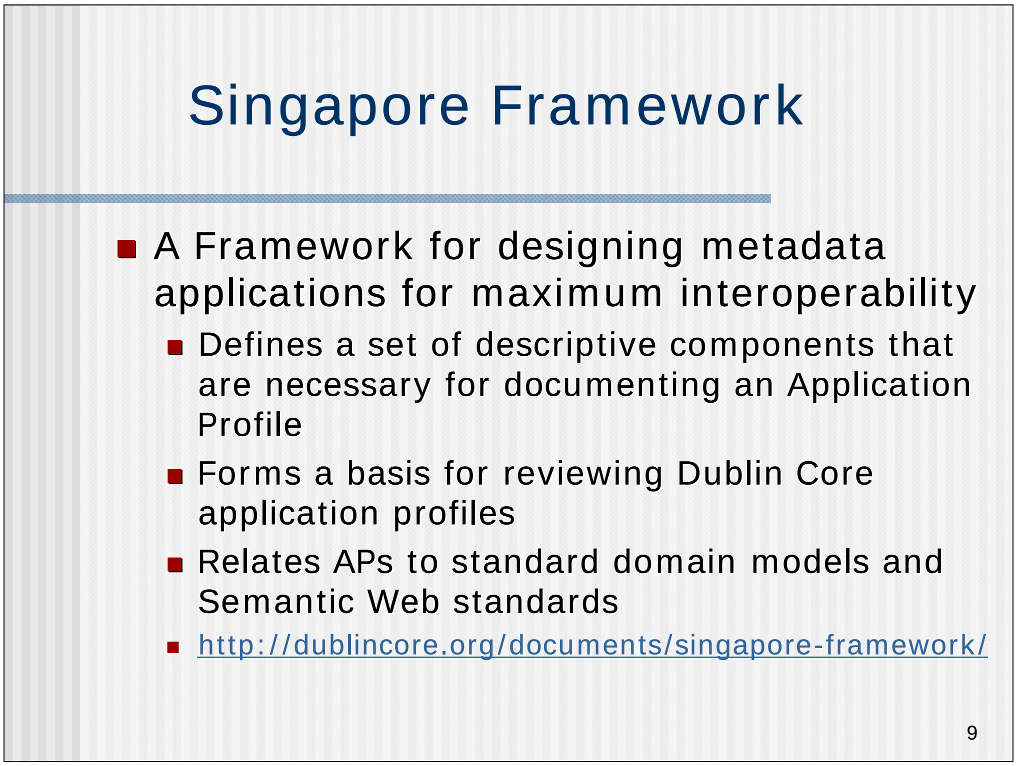
The debate about reuse of properties (i.e.data elements from another schema) centers around whether the additional properties use the same data model as Dublin Core. For instance, the Dublin Core Library Application Profile was originally written with several MODS elements, but the Dublin Core Usage Board rejected that approach, since MODS has so far not been modeled as RDF properties conforming to the Dublin Core Data Model (it is currently only in hierarchical XML). Work on modeling MODS as RDF has been done at MIT and is on the agenda for the MODS Editorial Committee.



Best practices for documenting a new element

Obligation to have a schema to back it up

Declare the schema to the registry



This work was a major output of the DCMI 2007 conference in Singapore and attempts to codify practices in creating Dublin Core Application Profiles.

Additional documentation about AP preparation has been developed, and work is progressing to finalize the framework.

DC Application Profile Examples

- Collections AP
  - http://www.dublincore.org/groups/collections/collection-application-profile/2007-03-09/
- Scholarly Works Application Profile (SWAP)
  - http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile
- Both these have been reviewed by the DC Usage Board and are deemed compliant with the DC Abstract Model

10

An RDA Application Profile

- A DCMI/RDA Task Group has been defining RDA properties and value vocabularies as formal RDF vocabularies (with URIs)
  - IFLA has stated an intention to declare FRBR entities and attributes as well
  - Next step is a DC application profile of RDA according to the Singapore Framework
  - See http://metadataregistry.org for the provisionally registered properties/vocabularies

11

METS Profiles

- 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)

12

METS also has the concept of application profiles (called just “profiles”).

Although not intended to be machine actionable, the standard syntax used can allow for validating a METS document for conformance to a given profile.

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

13

These are components that are defined in the formal METS profile schema.

Case study of a METS Profile

- LC Audio Compact Disc Profile
- Features:
  - Specifies MODS for descriptive metadata
  - Specifies description rules as AACR2
  - Specifies controlled vocabularies used in various elements
  - dmdSec requirements 2 and 3 specify use of relatedItem type="constituent" if there are multiple works on the CD
  - Specifies how to detail the physical structure, whether multiple CDs or multiple tracks on a CD (structMap requirements 2 and 3)

14

Link to the LC audio compact disc profile. The link is to the HTML display (the XML file is the same URL but with file extension "xml" instead of "html").

The profile is too detailed to go into in depth. Note the specification of controlled vocabularies used and how structural metadata is expressed in the METS structMap.

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

15

We are beginning to also see MODS profiles, again to specify specific rules, guidance, controlled vocabularies for particular applications.

DLF Aquifer has developed rich guidelines that institutions are starting to use as a metadata standard for exchange or harvest of descriptive metadata.

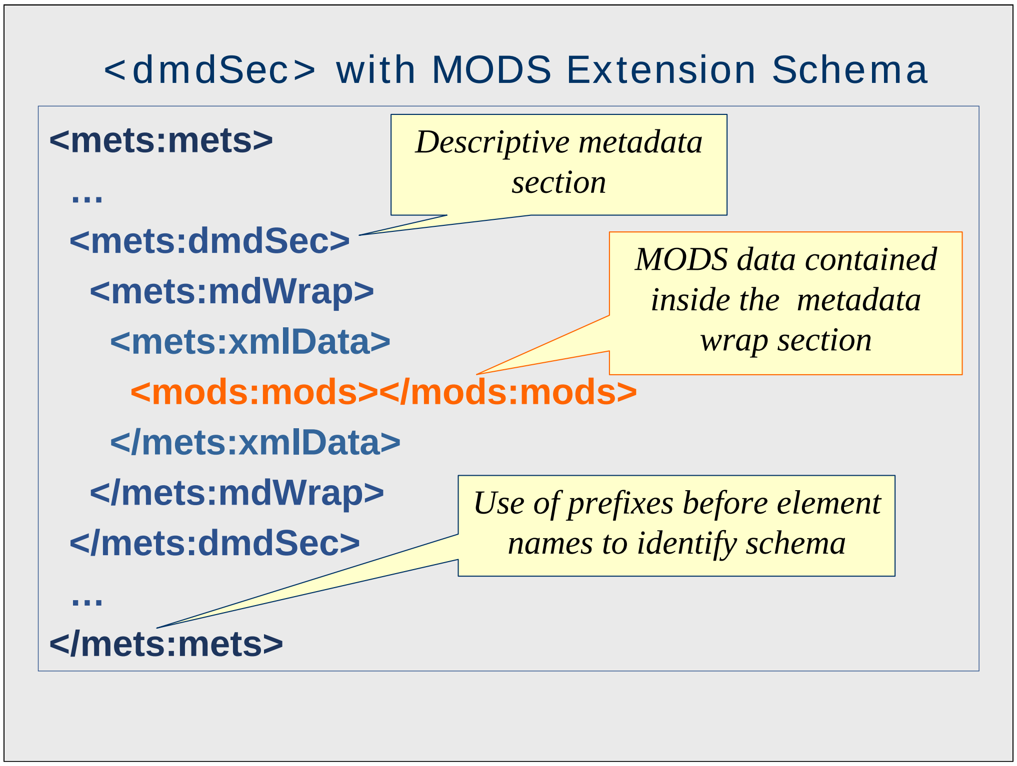
The BL electronic journal MODS profile is under development for a project involving legal deposit of ejournals using MODS for its descriptive metadata. (It also uses METS and PREMIS.)

Using metadata standards together: a case study

- METS can be used to package together the metadata with the objects
  - METS allows for use of any XML metadata schema in its extensions
  - 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
- METS can be used as a digital library application if objects are based on a profile and thus are consistent

16

We will look at how we can apply metadata standards in a digital library application using METS as the container. The METS container is used to bundle together the metadata as well as specify the files and their locations that are part of the entity. Objects based on a profile are consistent so that consistent behaviors can be applied, e.g. web presentations, page turning applications, etc.



This set of slides illustrates how METS and MODS work together in a digital library application. As we learned already, descriptive metadata may be inserted (“mdWrap”) in a METS descriptive metadata section. The use of prefixes shows the authority for the XML schema used and how two different namespaces are used in the same document.

```
<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>
```

Here is a simple illustration of a logical hierarchy created with relatedItems (using an audio CD). There are actually 3 levels of description: the item itself (the CD), a work contained within it (Symphony No. 5) and a movement within that work (Allegro con moto). The MODS hierarchy in relatedItem allows for this nested description.

This object uses the compact disc profile we just looked at, which specified how to detail the descriptive metadata of the components in the descriptive metadata section.

Use of MODS relatedItem *type="constituent"*

- A first level 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

19

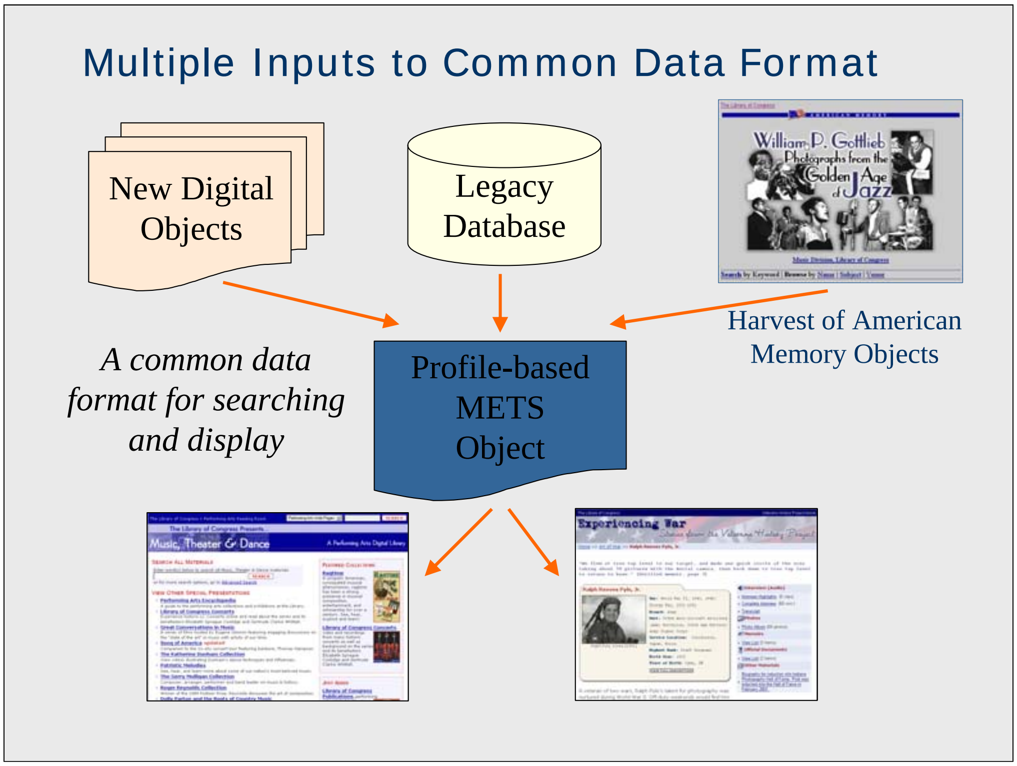
METS 2 Hierarchies: Logical & Physical

```
<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)

We can associate the physical items detailed in the METS structMap with any part of the hierarchy for the logical structure.



In the case of METS/MODS digital library projects, LC is taking metadata from different sources and using them in METS objects that are based on defined profiles. Examples are records from the ILS that are converted to MODS, non-MARC records from American Memory and records newly cataloged using MODS. This provides a common data format for searching and display. **Use of an application profile allows for consistent processing of the document in terms of descriptive metadata conversion, display and searching when bringing together metadata from different sources.**

Example: Using a profile as an application

- 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).

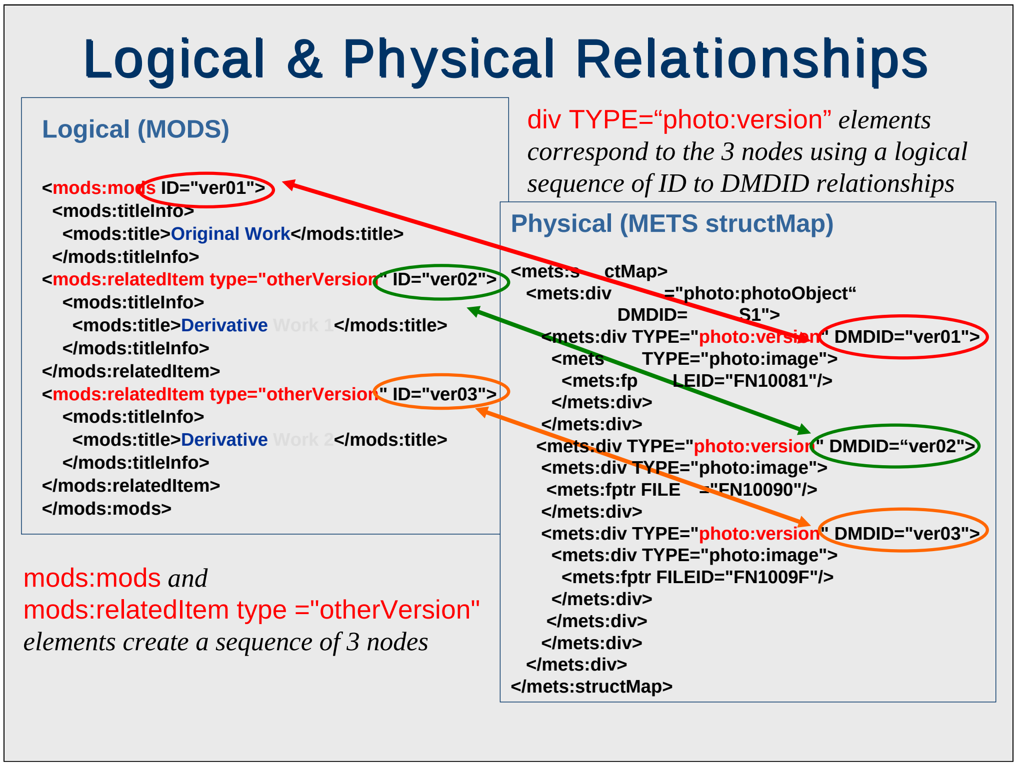
Used XSLT stylesheet to create 1600 records conforming to the METS photograph profile.

22

Here is an example of the use of a profile with MODS metadata and METS as a container format. It allows us to use METS not only as a container, but also as an application.

This uses the Photographic object profile. Records were converted from MARC to MODS. The profile allows for consistency and for the application of an XSLT stylesheet for presentation.

Link to this Portrait of Louis Armstrong and see how it is presented on the web. See (on the left of the screen) that there are 3 versions detailed in this METS document: the original negative, the contact print with annotations, and the published print.



The linking ID is used to show how the logical structure in the structural map relates to what is described in the MODS record. In this case there is an original and 3 derivative items (a common case in digitization projects). Information for each is given in MODS related item and related to the “div” in the structMap. The files for which the metadata is given is linked by the FILEID to the file section in the METS document.

Using a METS profile-based approach

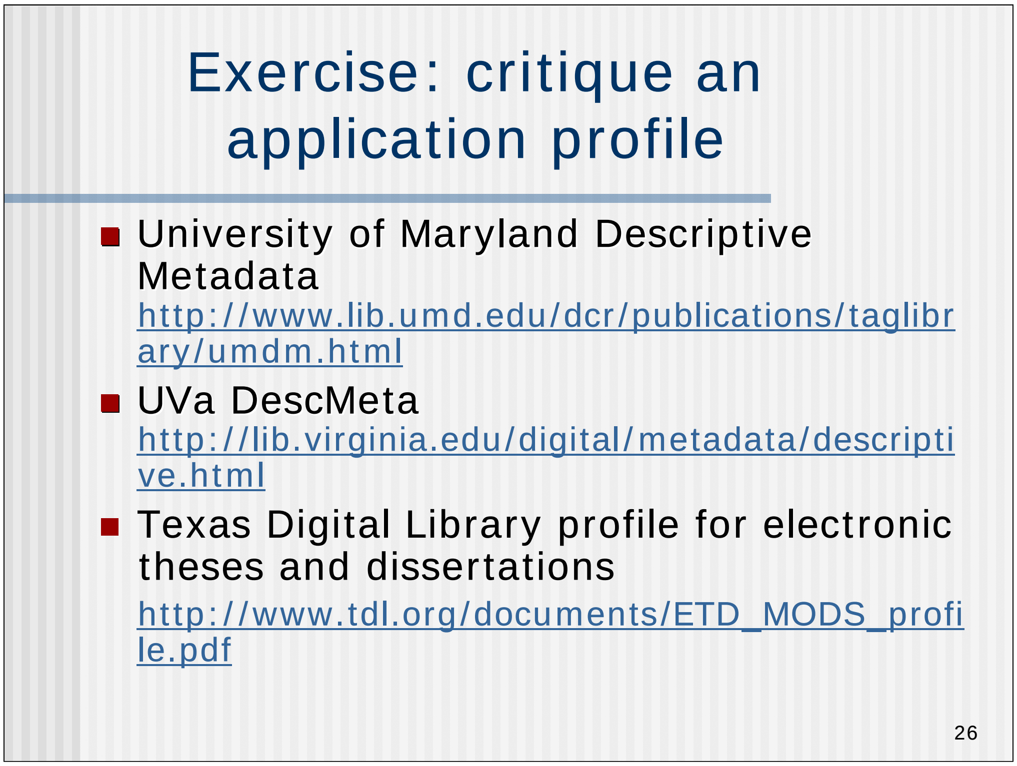
- Ability to model complex library objects
- Use of open source software tools
- Use of XML for data creation, editing, storage and searching
- Use of XSLT for...
  - Legacy data conversion
  - Batch METS creation and editing
  - Web displays and behaviors
- Creation of multiple outputs from XML
  - HTML/XHTML for Web display; PDF for printing
- Ability to aggregate disparate data sources into a common display

24

Closing thoughts on application profiles

- Many metadata standards are sufficiently flexible that profiling is necessary
- Documenting what is used in an application will simplify and enhance data presentation, conversion from other sources, ability to provide different outputs
- Constraining a metadata standard by specifying what is used and how facilitates data exchange and general interoperability

25



Divide into 3 groups to do this exercise. We will only look at descriptive metadata application profiles for this exercise.

Note to instructor: The link for the Texas digital library profile doesn’t work by clicking on it. You will need to copy and paste the URI into a web browser (it is the correct one) if you want to show it. Copies of this document are being provided in the student manuals.

Exercise: Questions to address

- Does the profile define its user community and expected uses?
- How usable would the profile be for a potential implementer?
- How (well) does the profile specify element/term usage?
- How (well) does the profile define and manage controlled vocabularies?
- Does the profile use existing metadata standards?
- Are there key anomalies, omissions, or implementation concerns?

27

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

University of Maryland Descriptive Metadata Tag Library

Authored by Jennifer O'Brien Roper Jennifer O'Brien Roper

Presentation designed by Sean Daugherty

This tag library provides current documentation for all descriptive metadata for digital objects in the University of Maryland Digital Collections. The University of Maryland Descriptive Metadata (UMDM) DTD and element set are based on Descriptive Metadata (Descmeta) DTD (Version 1.06) authored by Daniel McShane and Perry Roland at the University of Virginia. UMDM was designed within the local context, yet allows data to be adaptable to comply with national standards such as Dublin Core, VRA Core, or MODS.

The tag library contains definitions, notation of parent and child elements, attributes, input standards, and examples for each element. Wherever possible, links to appropriate external standards are present. UMDM is an evolving standard, and while stable, is not static. As new projects and material types are added to the repository, the standard may be modified.

- <extRef>
- <form>
- <geogName>
- <identifier>
- <imprint>
- <language>
- <linkGrp>
- <mediaType>
- <other>
- <persName>
- <physDesc>
- <pid>
- <price>
- <relation>
- <relationships>
- <repository>
- <rights>
- <series>
- <size>
- <style>
- <subject>
- <title>
- <version>

View Document Type Definition (DTD)

Required base elements:

Finite:

- <covPlace>
- <covTime>
- <mediaType>
- <physDesc>
- <pid>
- <relationships>
- <repository>
- <rights>

Repeatable:

- <culture>

- <description>
- <subject>
- <title>

Optional base elements:

Repeatable:

- <agent>
- <language>
- <style>
- <identifier>

<address>

Definition:

Contains a postal or other address, for example of an organization, individual, or internet resource. Addresses may be encoded either as a sequence of lines, or using any sequence of address component elements.

Must Contain:

Repeatable:

- <addressLine>

May Occur Within:

- <imprint>
- <repository>

Input Standard(s):

Address as found on item, or as found from a reliable source.

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta>
<relationships>
[...]
<relation type="isPartOf">Digital Collections @ UM</relation>
```

```
</relationships>
<repository>
<corpName>Historical Manuscripts</corpName>
<address>
<addressLine>http://www.lib.umd.edu/histmss/</addressLine>
</address>
</repository>
[...]
</descMeta>
```

<addressLine>

Definition:

Individual line of an address, postal or otherwise.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <address>

Input Standard(s):

Create as necessary to separate discrete portions of an address.

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta>
<relationships>
[...]
<relation type="isPartOf">Performing Arts Digital Videos</relation>
</relationships>
<repository>
<corpName>Nonprint Media Services</corpName>
<address>
<addressLine>Hornbake Library</addressLine>
<addressLine>College Park, MD 20742-7011</addressLine>
</address>
</repository>
[...]
</descMeta>
```

<agent>

Definition:

An entity primarily responsible for the creation or distribution of the intellectual content of a resource.

Must Contain At Least One of the Following:

Finite:

- <corpName>
- <other> *Use of this element is restricted. See the Usage Rules for Specific Collections for this element.*
- <persName>

May Occur Within:

- <bibRef>
- <descMeta>
- <imprint>

Input Standard(s):

Attribute type must be selected.

Attributes:

id	ID	Optional
label	CDATA	Optional
normal	CDATA	Optional
role	"author", "illustrator", "publisher", "broadcaster", "composer", "conductor", "director", "lyricist", "narrator", "performer", "producer", "speaker", "storyteller"	Optional
type	"creator", "contributor", "provider"	REQUIRED
xml:lang	CDATA	Optional

Specific Usage Rules for Collections

The following elements, attributes, and values may be used with this element in these specific conditions.

Prange Digital Children's Book Collection

- <agent type="creator" role="author">
- <agent type="contributor" role="illustrator">
- <agent type="provider" role="publisher">

Example:

```
<descMeta>
[...]
<title type="main">Agamemnon</title>
<agent type="creator">
<persName>Aeschylus</persName>
</agent>
<agent type="creator">
<persName>Harrison, Tony, 1937-</persName>
</agent>
<agent type="provider">
<corpName>National Theatre (Great Britain)</corpName>
</agent>
<agent type="provider">
<corpname>Channel Four (Great Britain)</corpName>
</agent>
<covPlace>
<geogName>Greece</geogName>
</covPlace>
[...]
</descMeta>
```

<availability>

Definition:

Supplies information about the availability of a resource.

Must Contain:

Repeatable:

- <date type=*[...]*>
- and/or <price>

May Occur Within:

- <imprint>

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta>
[...]
<relationships>
<relation label="collection" type="isPartOf" >Treature of World's Fair Art &
Architecture </relation>
<relation label="fair" type="isPartOf">Centennial Exhibition (1876: Philadelphia,
Pa.)</relation>
<relation type="isPartOf">
<bibRef>
<imprint>
<geogName>London</geogName>
<agent type="provider">
<corpName>John Murray</corpName>
</agent>
<address>
```

```
<addressLine>Albemarle-street</addressLine>
</address>
<date era="ad" type="exact">1818</date>
<availability>
<price units="shillings">10</price>
</availability>
</imprint>
</bibRef>
</relation>
</relationships>
[...]
</descMeta>
```

<bibRef>

Definition:

Bibliographic reference. Contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.

Must Contain:

Any of the following

Finite:

- <bibScope>
- <repository>
- <version>

Repeatable:

- <agent>
- <date type="*[...]*">
- <imprint type="*[...]*">
- <series>
- <title type="*[...]*">

May Occur Within:

- <corpName>
- <description>
- <linkGrp>
- <other>
- <persName>
- <relation>

Input Standard(s):

From the resource itself or an accepted authoritative source.

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Fixed
xml:lang	CDATA	Optional

Example:

```
<descMeta>
[...]
<physDesc>
<color>color</color>
<extent units="image">1</extent>
</physDesc>
<relationships>
<relation label="collection" type="isPartOf">University AlbUM</relation>
<relation label="citation" type="isPartOf">UM Call Number: ARCH REF GV 885.43.U535
U54 2004
<bibRef>Ungrady, Dave. Legends of Maryland basketball. Champaign, IL: Sports
Publishing, 2004.</bibRef>
</relation>
</relationships>
<repository>
<corpName>University Archives</corpName>
</repository>
[...]
</descMeta>
```

<bibScope>

Definition:

Scope of citation. Defines the scope of a bibliographic citation, for example as a list of page numbers, or a named subdivision of a larger work.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <bibRef>

Input Standard(s):

From the resource itself or an accepted authoritative source.

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta>
[...]
<relationships>
<relation type="isReferencedBy">
```

```
<bibRef>
<agent type="creator">
<persName>Earle Leighton, 1917-</persName>
</agent>
<title type="main">Confederate Broadside Verse</title>
<imprint>
<geogName>Texas</geogName>
<geogName>New Braunfels</geogName>
<agent type="provider">
<corpName>Book Farm</corpName>
</agent>
<date>1950</date>
</imprint>
<bibScope type="citation number">60</bibScope>
</bibRef>
</relation>
</relationships>
<repository>
<corpName>Art & Architecture Libraries</corpName>
</repository>
[...]
</descMeta>
```

<century>

Definition:

A numerical representation of the century or centuries associated with the coverage time or subject matter of a resource.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <covTime>
- <subject>

Input Standard(s):

YY01-YY00

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
normal	CDATA	Optional
certainty	"exact", "circa"	Optional
era	"ad", "bc", "cc", "cd"	REQUIRED
xml:lang	CDATA	Optional

Example:

```
<descMeta>
[...]
```

```
<covPlace>
<geogName>not captured</geogName>
</covPlace>
<covTime>
<century era="ad">1901-2000</century>
<dateRange era="ad" from="1974" to="1978">1974-1978</date>
</covTime>
<culture>American</culture>
<culture>Maryland</culture>
[...]
</descMeta>
```

<color>

Definition:

Describes the color quality of image and moving image resources.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <physDesc>

Input Standard(s):

Choose from the following list:

- black and white
- color
- monochrome

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta>
[...]
<identifier type="handle">hdl:1903.1/2240</identifier>
<physDesc>
<color>monochrome</color>
<extent units="image">1</size>
</physDesc>
<relationships>
<relation label="collection" type="isPartOf">University AlbUM</relation>
<relation label="archivalcollection" type="isPartOf">Records of the Department of
Intercollegiate Athletics</relation>
</relationships>
[...]
</descMeta>
```

<corpName>

Definition:

Corporate name. Contains the text for the name of a corporate entity.

May Contain:

Finite:

- PCDATA
- <bibRef>
- <extPtr>
- <extRef>
- <linkGrp>

May Occur Within:

- <agent>
- <repository>
- <subject>

Input Standard(s):

Library of Congress Name Authority File. For names not found in the LCNAF, enter corporate names in full direct form, omitting initial articles. Enter more than one agent if necessary for access.

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
normal	CDATA	Optional
role	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta xml:lang="en">
[...]
<agent type="contributor">
<persName>Price, Kenneth</persName>
</agent>
<agent type="contributor">
<persName>Cossons, Neil, 1939-</persName>
</agent>
<agent type="provider">
<corpName>HTV West (Firm)</corpName>
</agent>
<agent type="provider">
<corpName>Films for the Humanities (Firm)</corpName>
</agent>
[...]
</descMeta>
```

<covPlace>

Definition:

Coverage place. The geographical location associated with the production of the content of a resource (e.g. place of publication for a scanned image of a postcard). Provide information from broadest level (i.e. country or continent) to the most specific level known (region, state, city). Each location level should be provided in a separate <geogName> element.

Must Contain:

Repeatable:

- <geogName>

May Occur Within:

- <descMeta>

Input Standard(s):

Materials with an unknown place of origin may contain the phrase *"not captured"*. All other materials require place names encoded within <geogName> sub-elements.

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA, "printing"	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta xml:lang="en">
[...]
<covPlace>
<geogName type="continent">North America</geogName>
<geogName type="country">United States</geogName>
<geogName type="region">New Jersey</geogName>
</covPlace>
<covTime>
<century era="ad">1901-2000</century>
<date era="ad">1990</date>
</covTime>
[...]
</descMeta>
```

<covTime>

Definition:

Coverage time. Time period associated with the production of the content of the resource (e.g. date of publication for a video digitized from videotape). Identification of a century is required, and additional elements describing a specific date or date range are also available.

Must Contain:

Repeatable:

- <century>

May Contain:

Finite:

- <dateRange>
- <date>

May Occur Within:

- <descMeta>

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
xml:lang	CDATA	Optional

Specific Usage Rules for Collections

See the usage rules for the <date> element.

Example:

```
<descMeta xml:lang="en">
[...]
<covPlace>
<geogName type="continent">North America</geogName>
<geogName type="country">United States</geogName>
<geogName type="region">New Jersey</geogName>
</covPlace>
<covTime>
<century era="ad">1901-2000</century>
<date era="ad">1990</date>
</covTime>
[...]
</descMeta>
```

<culture>

Definition:

Contains text that indicates the culture of origin or context for a resource.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <descMeta>

Attributes:

id	ID	Optional
label	CDATA	Optional

type	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta xml:lang="en">
[...]
<covTime>
<century era="ad">1900-2000</century>
<date era="ad">1990</date>
</covTime>
<culture>European</culture>
<culture>British</culture>
<language>eng</language>
[...]
</descMeta>
```

<date>

Definition:

A date associated with an event in the life cycle of the resource. Most often, date will be associated with the creation or availability of the resource.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <availability>
- <bibRef>
- <covTime>
- <imprint>
- <rights>
- <subject>

Input Standard(s):

From the resource itself or an accepted authoritative source. Enter date in the form: YYYY-MM-DD. Range of dates may be expressed by using the ISO 8601 <dateRange> element. Enter more than one coverage-time if necessary for access.

Attributes:

id	ID	Optional
label	CDATA	Optional
type		Optional
normal	CDATA	Optional
certainty	exact, circa	Optional

era	"ad","bc","cc","cd"	REQUIRED
xml:lang	CDATA	Optional

Specific Usage Rules for Collections

The following elements, attributes, and values may be used for this element in these specific collections.

University AlbUM

- `<covTime><date certainty="exact">` *Required*: When using the <date> element in <covtime>, the *certainty* attribute is required with a value of "exact" or "circa".

Prange Digital Children's Book Collection

- `<covTime><date era="ad" label="pcbccd">` *Optional*

Example:

```
<descMeta>
[...]
<covPlace>
<geogName>not captured</geogName>
</covPlace>
<covTime>
<century era="ad">1901-2000</century>
<date era="ad" type="exact">1926-12-26</date>
</covTime>
<culture>Maryland</culture>
<culture>American</culture>
<culture>University of Maryland</culture>
[...]
</descMeta>
```

<dateRange>

Definition:

A range of dates associated with an event in the life cycle of the resource. Most often, a date range will be associated with the creation or availability of the resource.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <covTime>
- <rights>

Input Standard(s):

Enter dateRange in the form: YYYY-MM-DD to YYYY-MM-DD. Attributes from and to are required, following the same pattern.

Attributes:

id	ID	Optional

label	CDATA	Optional
type	CDATA	Optional
normal	"exact", "circa"	Optional
certainty	CDATA	Optional
era	"ad","bc","cc","cd"	REQUIRED
from	CDATA	REQUIRED
to	CDATA	REQUIRED
xml:lang	CDATA	Optional

Example:

```
<descMeta>
[...]
<covPlace>
<geogName>not captured</geogName>
</covPlace>
<covTime>
<century era="ad">1901-2000</century>
<dateRange era="ad" from="1931" to="1933">1930-1933</dateRange>
</covTime>
<culture>Maryland</culture>
<culture>American</culture>
<culture>University of Maryland</culture>
[...]
</descMeta>
```

<decade>

Definition:

A decade associated with the subject matter of a resource.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <subject>

Input Standard(s):

Enter decade in the form: YYY1-YYY0

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta>
[...]
```

```
<description type="summary">Aerial view of University of Maryland, College Park
campus at 1200 feet looking northwest.</description>
<subject scheme="TGM2" type="topical">Aerial views</subject>
<subject scheme="TGM1" type="topical">Buildings</subject>
<subject type="temporal">
<decade>1931-1940</decade>
</subject>
<subject type="geographical">
<geogName type="continent">North America</geogName>
<geogName type="country">United States</geogName>
<geogName type="region">Maryland</geogName>
<geogName type="settlement">College Park</geogName>
</subject>
[...]
</descMeta>
```

<descMeta>

Definition:

Descriptive metadata. Wrapper element for all elements used to describe a resource.

Must Contain:

Finite:

- <covTime>
- <mediaType>
- <pid>
- <relationships>
- <repository>

Repeatable:

- <covPlace>
- <culture>
- <description>
- <physDesc>
- <rights>
- <subject>
- <title type="*[...]*">

May Contain:

Repeatable:

- <agent>
- <identifier>
- <language>
- <style>

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
xml:lang	CDATA	Optional

Example:

The following elements constitute the minimum set of elements for a descMeta instance (i.e. those required by the DTD):

```
<descMeta>
<pid>[...]</pid>
<mediaType type="[...]">
[...]
<form>[...]</form>
</mediaType>
<title type="[...]>[...]</title >
<covPlace>
<geogName>[...]</geogName>
</covPlace>
<covTime>
<century>[...]</century>
</covTime>
<culture>[...]</culture>
<description>[...]</description>
<subject type="[...]>[...]</subject>
<physDesc>[...]</physDesc>
<relationships>
<relation type="[...]">[...]</relation>
</relationships>
<repository>
<corpName>[...]</corpName>
</repository>
<rights>[...]</rights>
</descMeta>
```

<description>

Definition:

A textual description of the content of the resource, including abstracts or summaries for document-like objects, content descriptions for visual resources, or other descriptions not included in other elements.

Must Contain:

Finite:

- PCDATA

Repeatable:

- <bibRef>
- <extPtr>
- <extRef>
- <linkGrp>

May Occur Within:

- <descMeta>
- <relation>

Input Standard(s):

Enter descriptive text, remarks, and comments about the object. This information can be taken from the item, or supplied by the agency if no structured description or abstract is available.

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA, "summary", "credits"	Optional
lang	CDATA	Optional
xml:lang	CDATA	Optional

Specific Usage Rules for Collections

The following elements, attributes and values may be used with this element in these specific collections.

Films@UM

- `<description type="summary">` *REQUIRED*
- `<description type="credits">` *Optional*

The Jim Henson Works

- `<description type="credits">` *Optional*
- `<description type="summary">` *REQUIRED*

Prange Digital Children's Books Collection

- `<description type="papertype" label="pcbcensor">` *Optional*
- `<description label="pcbcensor">` *Optional*
- `<description label="pcbnotes">` *Optional*

University AlbUM

- `<description type="summary">` *REQUIRED*

  World's Fair

  - `<description type="caption">` *Optional; used to denote a caption on an original item NOT a caption created to describe digital content*

Example:

```
<descMeta>
[...]
<covTime>
```

```
<century>1901-2000</century>
<date>1997</date>
</covTime>
<culture>European</culture>
<language>eng</language>
<description type="summary">Photo of the first official game played by the Baltimore
Orioles versus Jersey City in the original Baltimore Stadium after Oriole Park
burned down.</description>
[...]
</descMeta>
```

<extent>

Definition:

The number of a given measurement that comprises a resource (e.g. 3 images)

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <physDesc>

Input Standard(s):

Record the number of a given measurement for the whole of the resource being described.

Attribute units is required to indicate the unit of measure.

Attributes:

id	ID	Optional
label	CDATA	Optional
type	CDATA	Optional
units	CDATA, documents, image, pages, minutes, copies	REQUIRED
xml:lang	CDATA	Optional

Specific Usage Rules for Collections

The following elements, attributes, and values may be used with these elements in these specific collections.

Prange Digital Children's Book Collection

- <physDesc><extent units="documents" label="pcbcensor"> *Optional*

Example:

```
<descMeta xml:lang="en">
[...]
<physDesc>
<extent units="pages">12</extent>
</physDesc>
<physDesc>
<size units="papersize">B5</size>
```

```
</physDesc>
<physDesc>
<size units="cm">26 x 19</size>
</physDesc>
<physDesc type="format">Book</physDesc>
<physDesc type="format" xml:lang="ja-Hani">本</physDesc>
[...]
</descMeta>
```

<extPtr>

Definition:

Extended pointer. An empty linking element which connects the model to an external electronic object.

Must Contain:

EMPTY

May Occur Within:

- <corpName>
- <description>
- <extRef>
- <linkGrp>
- <other> *Use of this element is restricted. See the Usage Rules for Specific Collections for this element.*
- <persname>
- <relation>

Input Standard(s):

Use the attributes entityref to identify the external object.

Attributes:

behavior	CDATA	Optional
entityref	ENTITY	Optional
href	CDATA	Optional
id	ID	Optional
idref	IDREF	Optional
inline	"true"	
rel	"version" or "part"	Optional
role	CDATA	Optional
show	"embed", "new", "replace", "other", "none"	Optional

targettype	CDATA	Optional
title	CDATA	Optional
type	"simple"	Fixed
xlink	CDATA	Optional
xpointer	CDATA	Optional
xml:lang	CDATA	Optional

Example:

```
<descMeta>
[...]
<agent type="creator">
<corpName>London Weekend Television, ltd <extPtr type="simple" entityref="lwt"
title="Logo for LWT" show="embed" /></corpName>
</agent>
<agent type="creator">
<corpName>Reiner Moritz Associates</corpName>
</agent>
[...]
</descMeta>
```

<extRef>

Definition:

Extended reference. A linking element that can include text and subelements as part of a reference to an external electronic object.

May Contain:

Finite:

- PCDATA

Repeatable:

- <extPtr>

May Occur Within:

- <corpName>
- <description>
- <linkGrp>
- <other> *Use of this element is restricted. See the Usage Rules for Specific Collections for this element.*
- <persName>
- <relation>

Input Standard(s):

Use the attributes entityref or href to identify the external object.

Attributes:

behavior	CDATA	Optional

entityref	ENTITY	Optional
href	CDATA	Optional
id	ID	Optional
idref	IDREF	Optional
inline	"true"	
rel	"version" or "part"	Optional
role	CDATA	Optional
show	"embed", "new", "replace", "other", "none"	Optional
targettype	CDATA	Optional
title	CDATA	Optional
type	"simple"	Fixed
xlink	CDATA	Optional
xml:lang	CDATA	Optional
xpointer	CDATA	Optional

Example:

```
<descMeta>
[...]
<agent type="creator">
<corpName>
<extRef type="simple" href="http://www.londonweekend.com" title="ITv.com"
show="new">London Weekend Television, ltd </extRef>
</corpName>
</agent>
<agent type="creator">
<corpName>Reiner Moritz Associates</corpName>
</agent>
[...]
</descMeta>
```

<form>

Definition:

The form of the contents of the parent element, according to a non-Dublin Core scheme, e.g. poetry, prose, fiction, painting, sculpture, etc. The form element functions as a qualifier for the <mediaType> element.

Must Contain:

Finite:

- PCDATA

May Occur Within:

- <mediaType>

Content Standard(s):

- <mediaType> = **Collection**:
    - Collection
- <mediaType> = **Dataset**:
    - *To be determined*
- <mediaType> = **Image**:
    - Architecture
    - Decorative art
    - Drawing. Design. Illustration.
    - Painting
    - Photograph, artistic
    - Photograph, documentary
    - Print
    - Sculpture
- <mediaType> = **interactiveResource**:
    - Hypertext systems
    - Interactive video
- <mediaType> = **movingImage**:
    - *To be determined*
- <mediaType> = **software**:
    - *To be determined*
- <mediaType> = **sound**:
    - instrumental music
    - vocal music
    - instrumental and vocal music
    - electronic music
    - spoken word
    - nature or natural sounds
- <mediaType> = **text**:
    - Poetry
    - Drama

UVa Metadata Descriptive Elements (UVa DescMeta)

Metadata Home > UVa Metadata Descriptive Elements

UVa DescMeta is currently under development by the Metadata Steering Group at the University of Virginia Library. The element set, the minimal requirements, and the DTDs are still considered in-progress. The DTD will be released as UVa DescMeta version 2.0 when development and prototyping are complete.

DTD at: http://text.lib.virginia.edu/dtd/descmeta/descmeta.dtd
Documentation at: http://dl.lib.virginia.edu/html/descmeta/
(note: the documentation is NOT up to date with the current DTD. Please use with caution.)
Mappings at: http://www.lib.virginia.edu/digital/metadata/mappings.html

Descriptive Element Set

<agent>
<authority>
<covplace>
<covtime>
<culture>
<description>
<identifier>
<language>
<mediatype>
<mimetype>
<physdesc>
<pid>
<place>
<relationships>
<rights>
<style>
<subject>
<surrogate>
<time>
<title>

Minimal Requirements for Ingest into the UVa Digital Library

<pid>

- Surrogate identifier
- The Fedora PID
- “Unknown” is not an option

<title type="main" >

- Original title
- One and only one
- “Unknown” is not an option

<mediatype>

- Mediatype
- One and only one (not repeatable)
- “Unknown” is not an option

<rights type="access" >

- Access rights
- “Unknown” is not an option

<surrogate><time>

- The date the resource was put on the first server
- If unknown, may be populated with date of ingest into Fedora

<time>

- An original date of any date type
- In extreme situations, an appeal can be made to the MSG for populating the field with "Unknown";
  the MSG will decide if this is acceptable on a case-by-case basis.

Not required but highly recommended:

<agent>

- At least agent of any type
- Consider using "Unknown", if unavailable

Digital Initiatives
University of Virginia
PO Box 400112
Charlottesville, VA 22904-4112

© The Rector and Visitors of the University of Virginia

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

TDL Metadata Working Group

MODS Application Profile for Electronic Theses and Dissertations Version 1

December 2005

Texas Digital Library
Application Profile for Electronic Theses and Dissertations

Introduction

This MODS application profile for electronic theses and dissertations (ETDs) describes the best practices for descriptive metadata for members of the Texas Digital Library (TDL). This document defines the mandatory minimum elements for ETDs. Besides these elements, other valid MODS elements may be included in ETD records. Optional elements, subelements, and attributes are described throughout the document.

MODS elements for ETDs:

Title Information (mandatory)
Name of Author (mandatory)
Name of Thesis Advisor (mandatory)
Name of Committee Member (optional)
Name of Degree Grantor (mandatory)
Type of Resource (mandatory)
Genre (mandatory)
Origin Information (mandatory)
Language (mandatory)
Physical Description (mandatory)
Abstract (mandatory)
Subject (mandatory)
Identifier (mandatory)
Location (mandatory)
Degree Information (mandatory)
Record Information (mandatory)

Instructions for formatting and encoding:

Title Information

Mandatory practice: Encode the title information in a <mods:titleInfo> wrapper element. Encode the title proper in a <mods:title> subelement. Encode the subtitle in a <mods:subTitle> subelement.

Optional practice: Other valid subelements or attributes within the <mods:titleInfo> element may be used.

Example:

```
<mods:titleInfo>
   <mods:title>Critical processes and performance measures for patient safety
      systems in healthcare institutions
   </mods:title>
```

```
   <mods:subTitle>a Delphi study</mods:subTitle>
</mods:titleInfo>
```

Name of Author

Mandatory practice: Encode information about the name of the author in the <mods:name> wrapper element with the type attribute set to “personal.” Encode the MARC relator term “Author” in the <mods:roleTerm> subelement under the <mods:role> subelement. Encode the various parts of the name in the <mods:namePart> subelement. Include the type attribute in each <mods:namePart> subelement. The “given” and “family” name types are mandatory.

Optional practice: Encode the birthdate in a <mods:namePart> subelement with type set to “date”. Other valid subelements or attributes within the <mods:name> element may be used.

Example:

```
<mods:name type="personal">
   <mods:namePart type="given">Ralitsa B.</mods:namePart>
   <mods:namePart type="family">Akins</mods:namePart>
   <mods:namePart type="date">1967-</mods:namePart>
   <mods:role>
      <mods:roleTerm authority="marcrelator"
         type="text">Author</mods:roleTerm>
   </mods:role>
</mods:name>
```

Name of Thesis Advisor

Mandatory practice: Encode information about the thesis advisor in the <mods:name> wrapper element with the type attribute set to “personal.” Encode the MARC relator term “Thesis advisor” in the <mods:roleTerm> subelement under the <mods:role> subelement. Encode the various parts of the name in the <mods:namePart> subelement. Include the type attribute in each <mods:namePart> subelement. The “given” and “family” name types are mandatory. The element <mods:name> is repeatable for thesis advisors.

Optional practice: Encode the birthdate in a <mods:namePart> subelement with type set to “date”. Other valid subelements or attributes within the <mods:name> element may be used.

Example:

```
<mods:name type="personal">
   <mods:namePart type="given">Bryan R.</mods:namePart>
   <mods:namePart type="family">Cole</mods:namePart>
   <mods:role>
      <mods:roleTerm authority="marcrelator" type="text">Thesis
         advisor</mods:roleTerm>
   </mods:role>
</mods:name>
```

Name of Committee Member

Optional practice: Encode information about committee members in the <mods:name> wrapper element with the type attribute set to “personal.” Encode the term “Committee member” in the <mods:roleTerm> subelement under the <mods:role> subelement. Encode the various parts of the name in the <mods:namePart> subelement. Include the type attribute in each <mods:namePart> subelement. The “given” and “family” name types are mandatory. The element <mods:name> is repeatable for committee members. Encode the birthdate in a <mods:namePart> subelement with type set to “date”. Other valid subelements or attributes within the <mods:name> element may be used.

Example:

```
<mods:name type="personal">
   <mods:namePart type="given">Jane R.</mods:namePart>
   <mods:namePart type="family">Smith</mods:namePart>
   <mods:role>
      <mods:roleTerm type="text">Committee member</mods:roleTerm>
   </mods:role>
</mods:name>
```

Name of Degree Grantor

Mandatory practice: Encode information about the degree grantor in the <mods:name> wrapper element with the type attribute set to “corporate.” Encode the name of the degree granting institution in a <mods:namePart> subelement. Use the form of the name authorized by the Library of Congress Name Authority File. Encode the name of the department that granted the degree in a <mods:namePart> subelement. Encode the MARC relator term “Degree grantor” in the <mods:roleTerm> subelement under the <mods:role> subelement.

Optional practice: Other valid subelements or attributes within the <mods:name> element may be used.

Example:

```
<mods:name type="corporate" authority="lcnaf">
   <mods:namePart>Texas A &amp; M University</mods:namePart>
   <mods:namePart>Philosophy</mods:namePart>
   <mods:role>
      <mods:roleTerm authority="marcrelator" type="text">
        Degree grantor
      </mods:roleTerm>
   </mods:role>
</mods:name>
```

Type of Resource

Mandatory practice: Encode the type of resource in the <mods:typeOfResource> element. The element <mods:typeOfResource> is repeatable for ETDs with multiple files.

Example:

```
<mods:typeOfResource>
   text
</mods:typeOfResource>
```

Genre

Mandatory practice: Encode the MARC genre term "theses" in the <mods:genre> element. Set the authority attribute to "marcgt."

Optional practice: Other valid attributes within the <mods:genre> element may be used.

Example:

```
<mods:genre authority="marcgt">
   theses
</mods:genre>
```

Origin Information

Mandatory practice: Encode relevant dates for the ETD in the <mods:originInfo> wrapper element. The creation date is defined as the date the student graduates or the date the degree is conferred. The publication date is defined as the date the ETD is released to the public.

Encode the month and year of the creation date, according to ISO 8601, in the <mods:dateCreated> subelement. Set the encoding attribute to "iso8601."

Encode the month and year of the publication date, according to ISO 8601, in the <mods:dateIssued> subelement. Set the encoding attribute to "iso8601."

Optional practice: The day of the month may be included date encodings. Other valid subelements or attributes within the <mods:originInfo> element may be used.

Examples:

```
<mods:originInfo>
   <mods:dateCreated encoding="iso8601">200408</mods:dateCreated>
   <mods:dateIssued encoding="iso8601">200412</mods:dateIssued>
</mods:originInfo>
```

Language

Mandatory practice: Encode information about the language of the ETD in the <mods:language> wrapper element. Encode the language, according to ISO 639-2b, in the <mods:languageTerm> subelement. Set the type attribute to "code" and the authority attribute to "iso639-2b." The <mods:languageTerm> subelement is repeatable.

Optional practice: Other valid subelements or attributes within the <mods:language> element may be used.

Example:

```
<mods:language>
   <mods:languageTerm type="code" authority="iso639-2b">
      eng
   </mods:languageTerm>
   <mods:languageTerm type="code" authority="iso639-2b">
      spa
   </mods:languageTerm>
</mods:language>
```

Physical Description

Mandatory practice: Encode the physical description in the <mods:physicalDescription> wrapper element. Encode the MARC format term "electronic" in the <mods:form> subelement. Set the authority attribute to "marcform." Encode the MIME type in the <mods:internetMediaType> subelement. Encode the digital origin ("born digital" or "reformatted digital") in the <mods:digitalOrigin> subelement.

Optional practice: Other valid attributes within the <mods:physicalDescription> element may be used.

Example:

```
<mods:physicalDescription>
   <mods:form authority="marcform">electronic</mods:form>
   <mods:internetMediaType>application/pdf</mods:internetMediaType>
   <mods:digitalOrigin>born digital</mods:digitalOrigin>
</mods:physicalDescription>
```

Abstract

Mandatory practice: Encode the abstract in the <mods:abstract> element. Include the language attribute encoded in ISO 639-2b.

Optional practice: Valid attributes within the <mods:abstract> element may be used.

Example:

```
<mods:abstract lang="eng">This dissertation study presents a conceptual
   framework for implementing and assessing patient safety systems in
   healthcare institutions. The conceptual framework consists of critical
   processes and performance measures identified in the context of the 2003
   Malcolm Baldrige National Quality Award (MBNQA) Health Care Criteria for
   Performance Excellence…</mods:abstract>
```

Subject

Mandatory practice: Encode topical subject terms in the <mods:subject> wrapper element. Encode individual terms or phrases in the <mods:topic> subelement. The <mods:subject> element is repeatable.

Optional practice: Controlled subject headings may be included by using the authority attribute of the <mods:topic> subelement. Other valid subelements or attributes within the <mods:subject> element may be used.

Example:

```
<mods:subject>
   <mods:topic>healthcare</mods:topic>
</mods:subject>

<mods:subject>
   <mods:topic>patient safety</mods:topic>
</mods:subject>

<mods:subject authority="lcsh">
   <mods:topic>Medical care</mods:topic>
   <mods:topic>Quality control</mods:topic>
   <mods:geographic>United States</mods:geographic>
   <mods:temporal>20th century</mods:temporal>
</mods:subject>
```

Identifier

Mandatory practice: Encode the unique identifier in the <mods:identifier> element. The <mods:identifier> element is repeatable.

Optional practice: The type attribute may be used in the <mods:identifier> element. Other valid attributes within the <mods:identifier> element may be used.

Example:

```
<mods:identifier type="hdl">
```

```
   http://handle.tamu.edu/1969.1/1042
</mods:identifier>
```

Location

Mandatory practice: Encode the location in the <mods:location> wrapper element. Encode the uniform resource locator (URL) in the <mods:url> subelement.

Optional practice: Other valid attributes within the <mods:identifier> element may be used.

Example:

```
<mods:location>
   <mods:url>
      http://handle.tamu.edu/1969.1/1042
   </mods:url>
</mods:location>
```

Degree Information

Note: The MODS standard does not have elements specifically for theses and dissertations. In order to encode degree information in MODS, the <mods:extension> element is used to reference the ETD-MS XML schema.

Mandatory practice: Encode information about the conferred degree in the <etd:degree> wrapper element. Encode the degree name in the <etd:name> subelement. Use the fully spelled out form of the degree name. Encode the degree level, from the TDL vocabulary, in the <etd:level> subelement. Encode the degree discipline, from the TDL vocabulary, in the <etd:discipline> subelement.

Example:

```
<mods:extension>
   <etd:degree>
      <etd:name>Doctor of Philosophy</etd:name>
      <etd:level>Doctoral</etd:level>
      <etd:discipline>Educational Administration</etd:discipline>
   </etd:degree>
</mods:extension>
```

Record Information

Mandatory practice: Encode information about the MODS record in the <mods:recordInfo> wrapper element. Encode the name of the agency that created the MODS record in the <mods:recordContentSouce> subelement, with the authority attribute set to “marcorg.” Encode the month, year, and day of the creation date of the record, according to ISO 8601, in the

<mods:recordCreationDate> subelement. Set the encoding attribute to “iso8601.” Encode the month, year, and day of the change date, according to ISO 8601, in the <mods:recordChangeDate> subelement. Set the encoding attribute to “iso8601.” Encode the unique record identifier in the <mods:recordIdentifier> subelement.

Optional practice: Other valid attributes within the <mods:recordInfo> element may be used.

Example:

```
<mods:recordInfo>
   <mods:recordContentSource authority="marcorg">
      TxCM
   </mods:recordContentSource>
   <mods:recordCreationDate encoding="iso8601">
      20050826
   </mods:recordCreationDate>
   <mods:recordChangeDate encoding="iso8601">
      20050826
   </mods:recordChangeDate>
   <mods:recordIdentifier>12345678</mods:recordIdentifier>
</mods:recordInfo>
```

Full example of MODS record for an ETD

```
<?xml version="1.0" encoding="UTF-8"?>
<mods:mods

xmlns:mods="http://www.loc.gov/mods/v3"
xmlns:etd="http://www.ndltd.org/standards/metadata/etdms/1.0/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation=
"http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-1.xsd

http://www.ndltd.org/standards/metadata/etdms/1.0/
http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd">

<mods:titleInfo lang="eng">
   <mods:title>Critical processes and performance measures for patient safety
      systems in healthcare institutions</mods:title>
   <mods:subTitle>a Delphi study</mods:subTitle>
</mods:titleInfo>

<mods:name type="personal" authority="lcnaf">
   <mods:namePart>Akins, Ralitsa B., 1967-</mods:namePart>
   <mods:namePart type="given">Ralitsa B.</mods:namePart>
   <mods:namePart type="family">Akins</mods:namePart>
   <mods:namePart type="date">1967-</mods:namePart>
   <mods:role>
      <mods:roleTerm authority="marcrelator"
         type="text">Author</mods:roleTerm>
   </mods:role>
</mods:name>

<mods:name type="personal">
   <mods:namePart type="given">Bryan R.</mods:namePart>
   <mods:namePart type="family">Cole</mods:namePart>
   <mods:role>
      <mods:roleTerm authority="marcrelator" type="text">Thesis
         advisor</mods:roleTerm>
   </mods:role>
</mods:name>

<mods:name type="corporate" authority="lcnaf">
   <mods:namePart>Texas A &amp; M University</mods:namePart>
   <mods:role>
      <mods:roleTerm authority="marcrelator" type="text">Degree
         grantor</mods:roleTerm>
   </mods:role>
</mods:name>

<mods:typeOfResource>
   text
</mods:typeOfResource>

<mods:genre authority="marcgt">
   theses
</mods:genre>
```

```xml
<mods:originInfo>
   <mods:dateCreated encoding="iso8601">200408</mods:dateCreated>
   <mods:dateIssued encoding="iso8601">200411</mods:dateIssued>
</mods:originInfo>

<mods:language>
   <mods:languageTerm type="code" authority="iso639-2b">
      eng
   </mods:languageTerm>
</mods:language>

<mods:physicalDescription>
   <mods:form authority="marcform">electronic</mods:form>
   <mods:internetMediaType>application/pdf</mods:internetMediaType>
   <mods:digitalOrigin>born digital</mods:digitalOrigin>
</mods:physicalDescription>

<mods:abstract lang="eng">This dissertation study presents a conceptual
   framework for implementing and assessing patient safety systems in
   healthcare institutions. The conceptual framework consists of critical
   processes and performance measures identified in the context of the 2003
   Malcolm Baldridge National Quality Award (MBNQA) Health Care Criteria for
   Performance Excellence. Methodology: The Delphi technique for gaining
   consensus from a group of experts and forecasting significant issues in
   the field of the Delphi panel expertise was used. Data collection included
   a series of questionnaires where the first round questionnaire was based
   on literature review and the MBNQA criteria for excellence in healthcare,
   and tested by an instrument review panel of experts. Twenty-three experts
   (MBNQA healthcare reviewers and senior healthcare administrators from
   quality award winning institutions) representing 18 states participated in
   the survey rounds. The study answered three research questions: (1) What
   are the critical processes that should be included in healthcare patient
   safety systems? (2) What are the performance measures that can serve as
   indicators of quality for the processes critical for ensuring patient
   safety? (3) What processes will be critical for patient safety in the
   future? The identified patient safety framework was further transformed
   into a patient safety tool with three levels: basic, intermediate, and
   advanced. Additionally, the panel of experts identified the major barriers
   to the implementation of patient safety systems in healthcare
   institutions. The identified "top seven" barriers were directly related to
   critical processes and performance measures identified as "important" or
   "very important" for patient safety systems in the present and in the
   future. This dissertation study is significant because the results are
   expected to assist healthcare institutions seeking to develop high quality
   patient safety programs, processes and services. The identified critical
   processes and performance measures can serve as a means of evaluating
   existing patient safety initiatives and guiding the strategic planning of
   new safety processes. The framework for patient safety systems utilizes a
   systems approach and will support healthcare senior administrators in
   achieving and sustaining improvement results. The identified patient
   safety framework will also assist healthcare institutions in using the
   MBNQA Health Care Criteria for Performance Excellence for self-assessment
   and quality improvement.</mods:abstract>

<mods:subject>
   <mods:topic>healthcare</mods:topic>
```

```
</mods:subject>

<mods:subject>
   <mods:topic>patient safety</mods:topic>
</mods:subject>

<mods:subject authority="lcsh">
   <mods:topic>Medical care</mods:topic>
   <mods:topic>Quality control</mods:topic>
   <mods:geographic>United States</mods:geographic>
   <mods:temporal>20th century</mods:temporal>
</mods:subject>

<mods:identifier type="hdl">
   http://handle.tamu.edu/1969.1/1042
</mods:identifier>

<mods:location>
   http://handle.tamu.edu/1969.1/1042
</mods:location>

<mods:extension>
   <etd:degree>
      <etd:name>Ph. D.</etd:name>
      <etd:level>Doctoral</etd:level>
      <etd:discipline>Educational Administration</etd:discipline>
   </etd:degree>
</mods:extension>

<mods:recordInfo>
   <mods:recordContentSource authority="marcorg">
      TxCM
   </mods:recordContentSource>
   <mods:recordCreationDate encoding="iso8601">
      20050826
   </mods:recordCreationDate>
   <mods: recordChangeDate encoding="iso8601">
      20050826
   </mods:recordChangeDate>
   <mods:recordIdentifier>12345678</mods:recordIdentifier>
</mods:recordInfo>

</mods:mods>
```

6. Controlled vocabularies

Metadata Standards and Applications Workshop

Goals of Session

- Understand how different controlled vocabularies are used in metadata
- Learn about relationships between terms in thesauri
- Understand methods of encoding vocabularies
- Learn about how registries are used to document vocabularies

2

Why controlled vocabularies?

- Document values that occur in metadata
- Goal is to reduce ambiguity
- Allow for control of synonyms
- Establish formal relationships among terms (where appropriate)
- Test and validate terms
- Role of metadata registries

3

Many metadata schemes allow for content from other sources. Some data elements may be more useful if a controlled vocabulary is used. Some are published formally, others are developed and used locally.

Formal controlled vocabularies may be used for testing and validation of terms– this is often done in integrated library systems, where bibliographic records may validate against authority records. This is one instance of testing and validation of terms. There is work being done on establishing metadata registries for both documentation and machine validation of both controlled vocabularies and metadata elements/terms. This could be particularly useful for controlled vocabularies, since their usefulness depends on consistency.

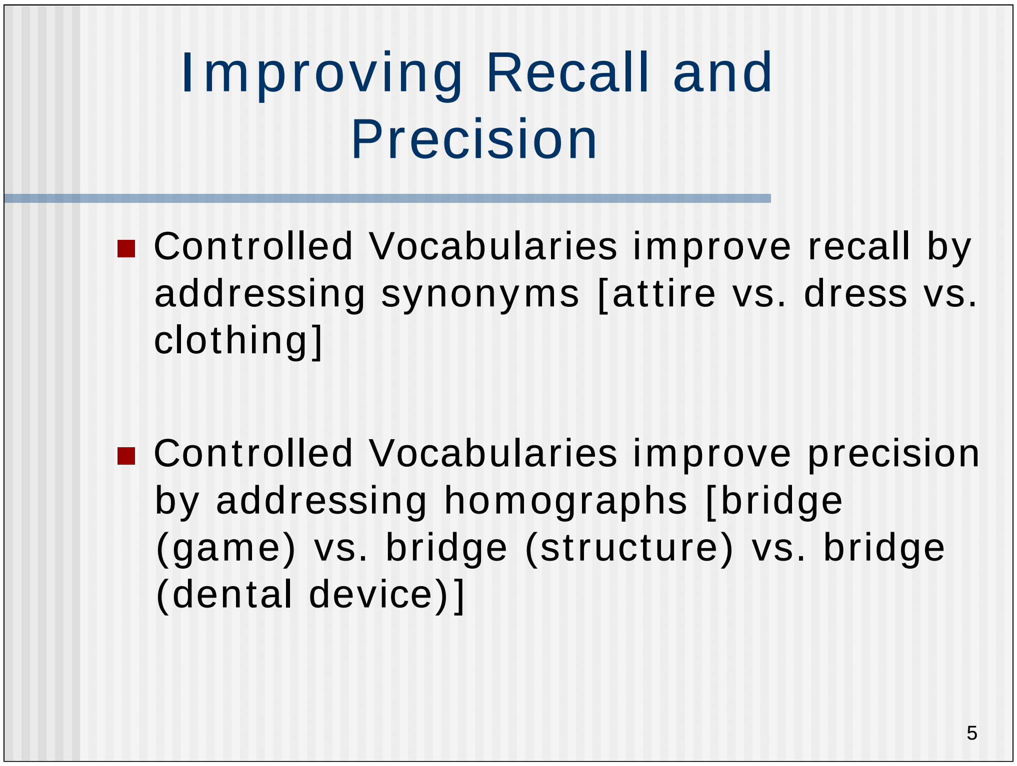
Why bother?

- 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

Research in library science shows that precision and recall ARE important to people

Precision: how much is noise and garbage?



All these “bridges” are very different things!

Types of Controlled Vocabularies

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

6

NISO has a standard for constructing thesauri (free download – in bibliography)

ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

With these kind of lists, the user is assumed to know what the values mean, or close enough.

There are also cases where enumerated lists are used in metadata schemes, perhaps in XML schemas. It would be useful if definitions were available for such values. An example is the PREMIS data dictionary, which has suggested value lists in the descriptions of the elements that take a controlled value.

Synonym Rings

- Synonym rings are used to expand queries for content objects
  - If a user enters any one of these terms as a query to the system, all items are retrieved that contain any of the terms in the cluster
- Synonym rings are often used in systems where the underlying content objects are left in their unstructured natural language format
  - the control is achieved through the interface by drawing together similar terms into these clusters
- Synonym rings are used in conjunction with search engines and provide a minimal amount of control of the diversity of the language found in the texts of the underlying documents

8

Unstructured: for example, full text searching.

Attempts to deal with the evolution of language over time

Taxonomies

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

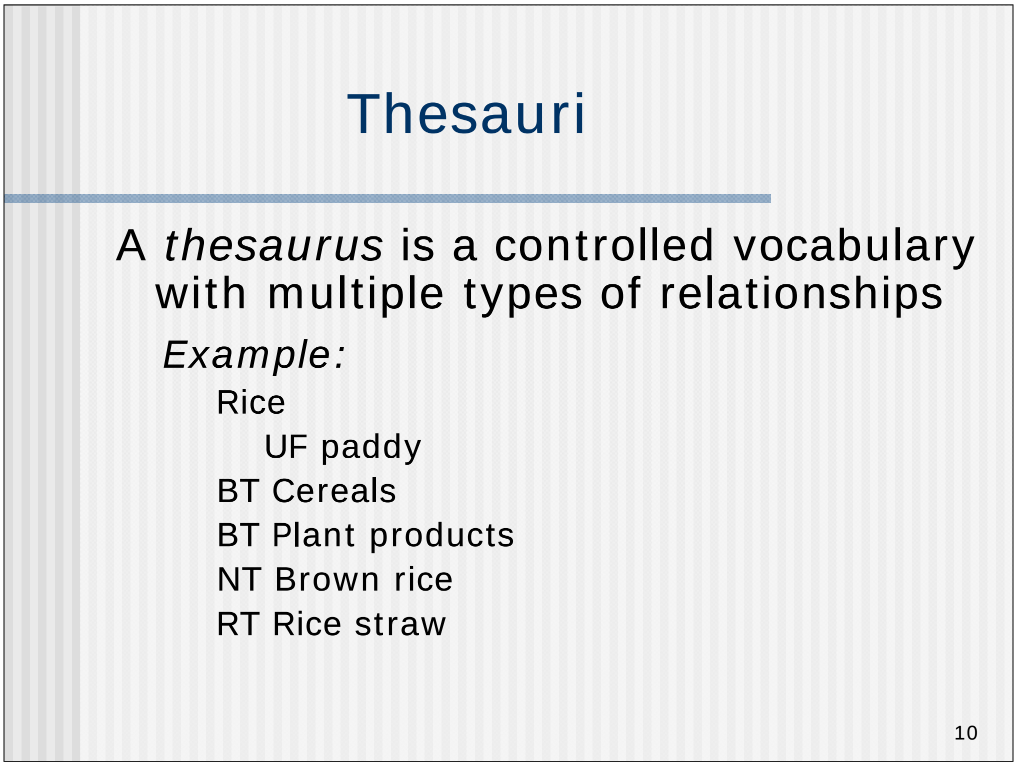
Example:

Chemistry
Organic chemistry
Polymer chemistry
Nylon

Frequently used in web navigation systems

9

“Nylon” could appear in other hierarchies as well, for example taxonomies of biological terms



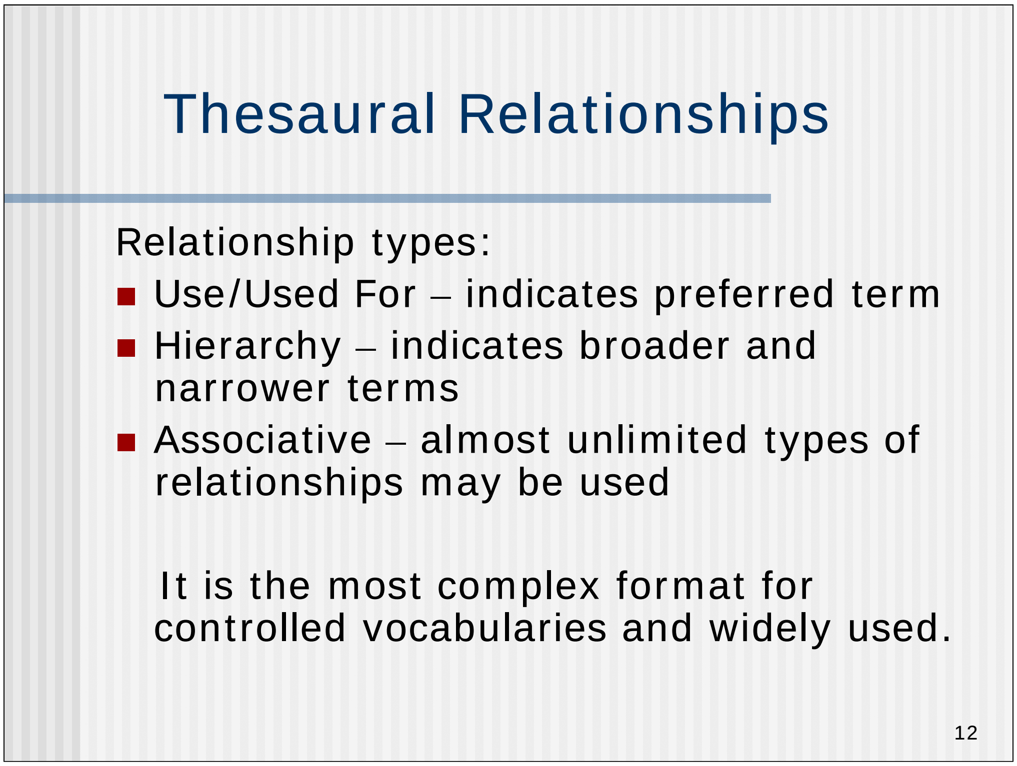
LCSH could be considered a thesaurus.

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

11

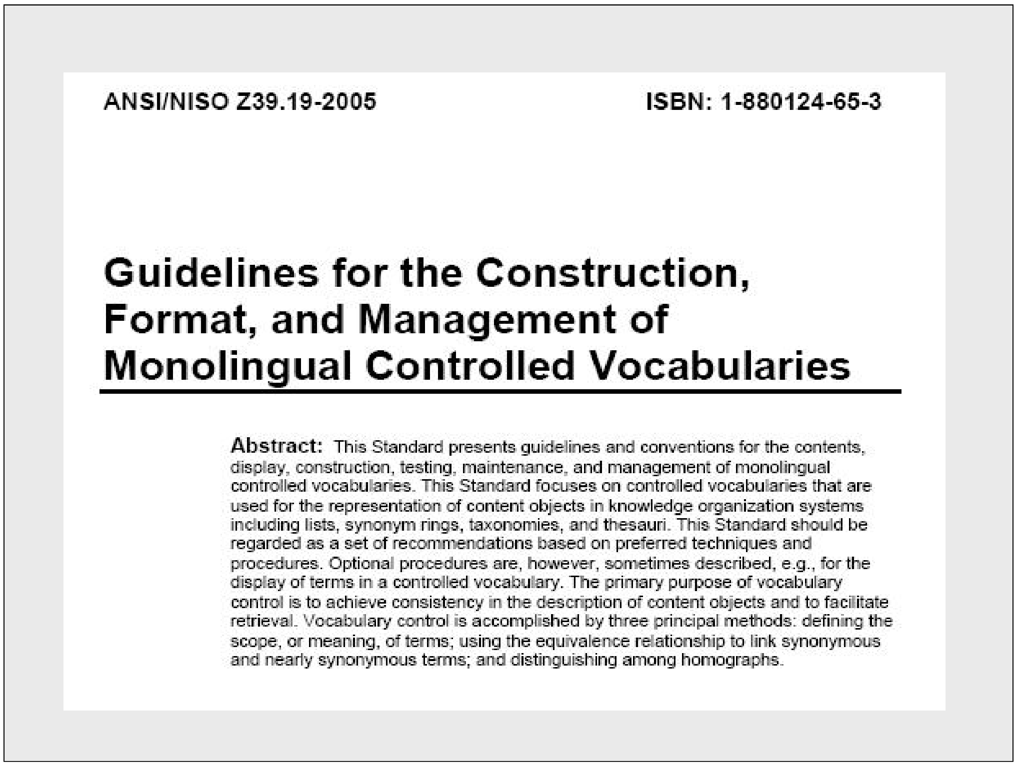
Ontology: this word has been used and misused.



Thesauri are the most useful, and work is being done in making them more useful in terms of coding.

A hierarchy lets you build many different levels.

Associative relationships can be expressed in different ways.



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.

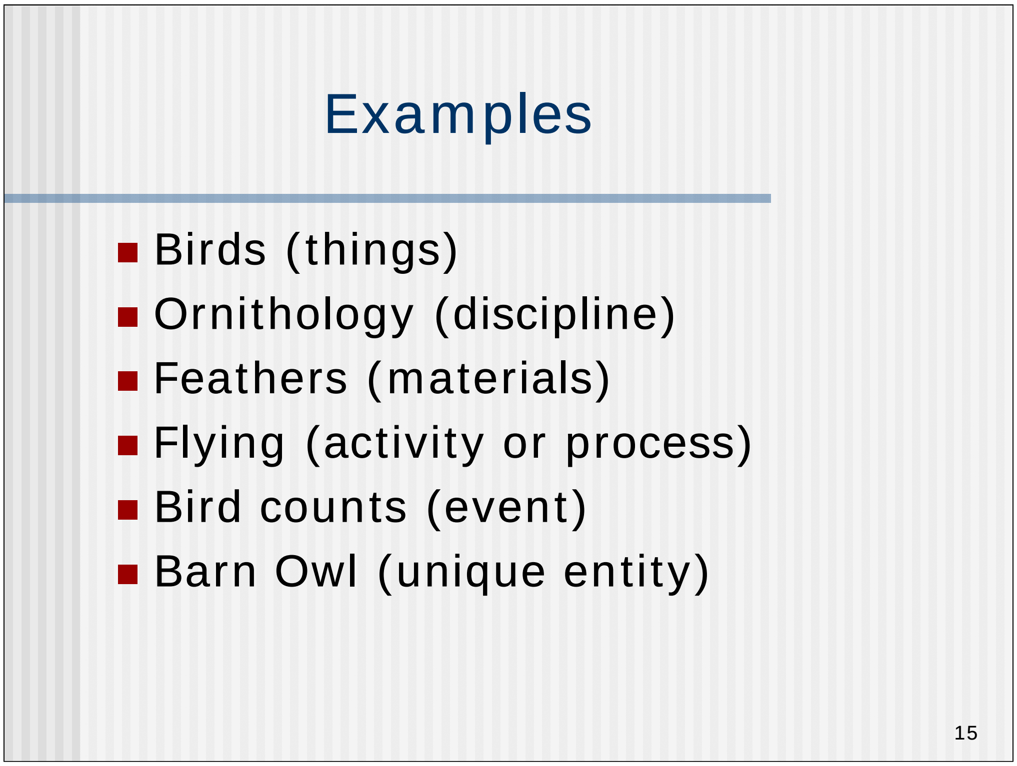
Note that Z39.19 is the content standard for establishing controlled vocabularies (it is independent of syntax used for expressing them).

Z39.19 Types of Concepts

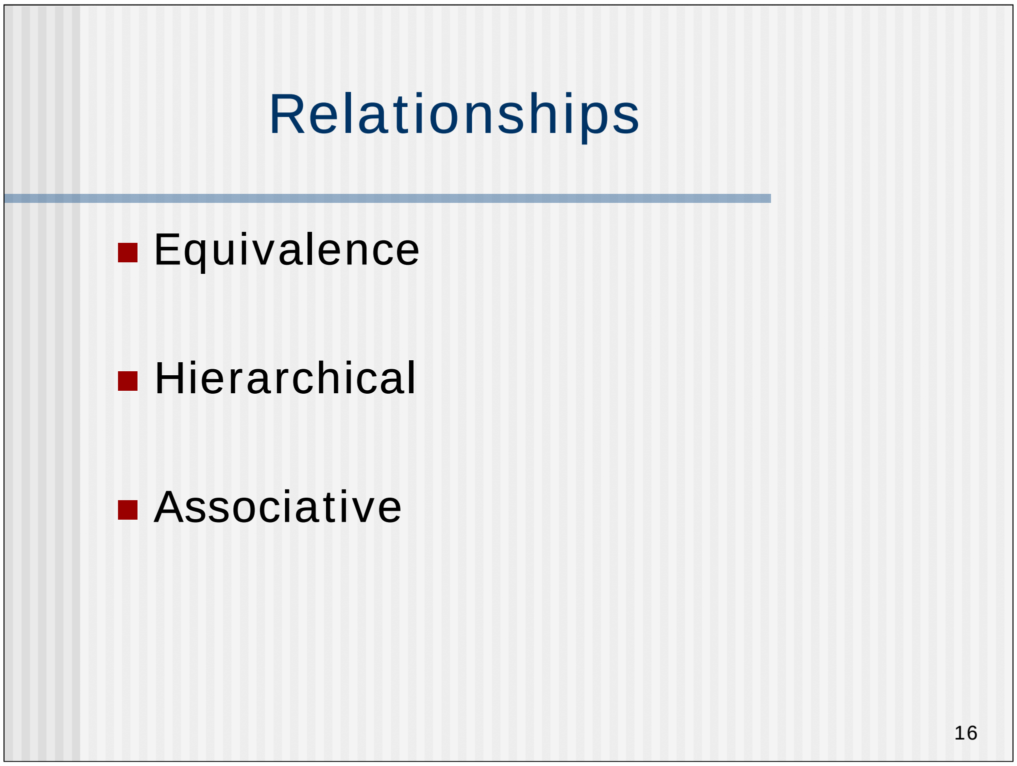
- Things and their physical parts
- Materials
- Activities or processes
- Events or occurrences
- Properties or states of persons, things, materials or actions
- Disciplines or subject fields
- Units of measurement
- Unique entities

14

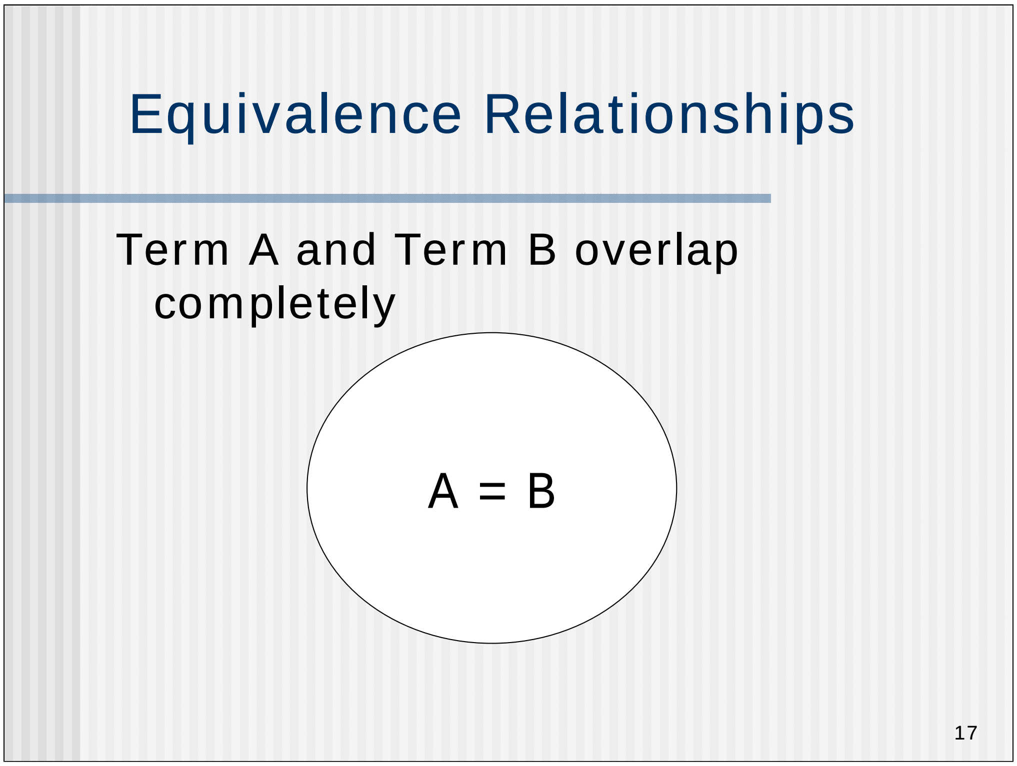
Concepts are categorized in this way because of the importance of associating things.



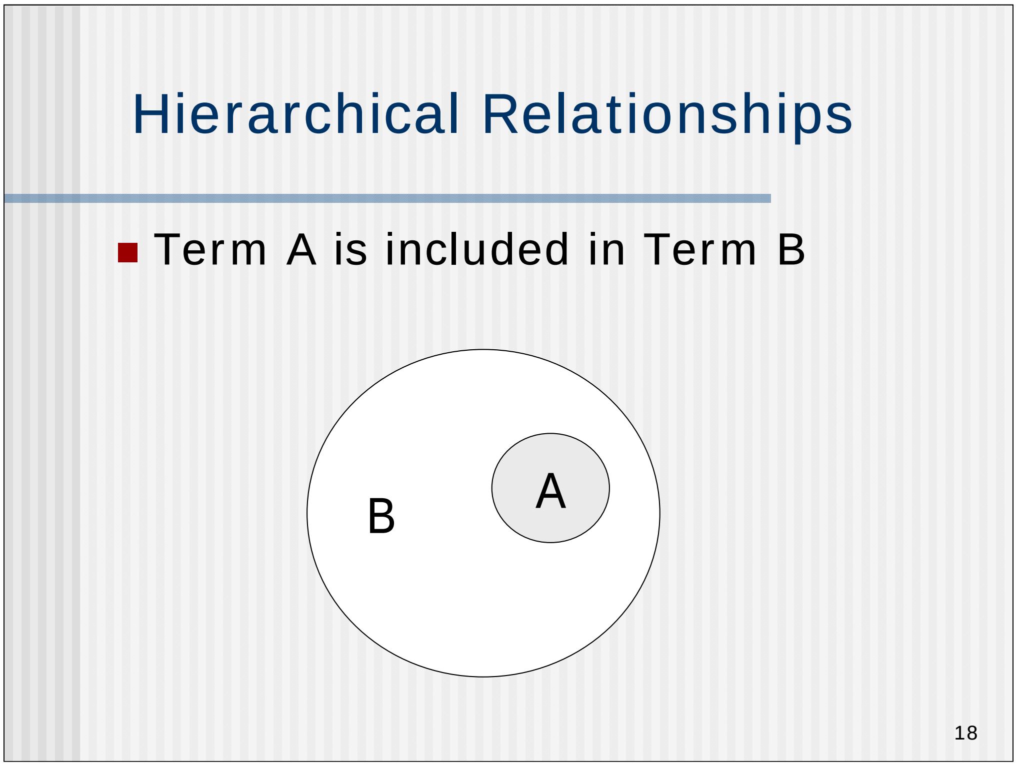
These examples are drawn from birds.



This is the big piece of any thesaurus.



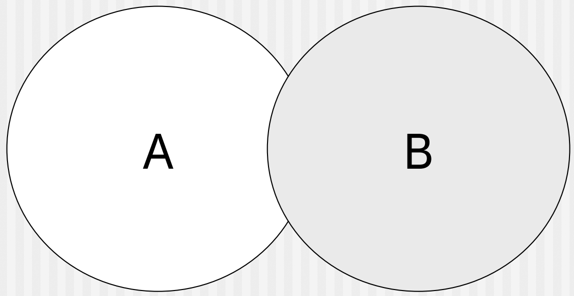
We will review some of the relationships that are possible between controlled terms.



One is broader; the other is narrower.

Associative Relationships

- Semantics of terms A and B overlap



19

Semantics are the meaning– associative relationships partially overlap.

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

20

Thesauri have conventions for expressing relationships.

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

21

Formal expression of vocabularies

Not folksonomies, not tagging

If you control a vocabulary, the degree of control depends on what kind of vocabulary it is.

A controlled vocabulary is based on a documented process.

Informal Vocabularies

- New movement towards 'bottom up' classification goes by many names:
  - Tagging
  - Social bookmarking
  - Folksonomies
- Some in this movement, seeing problems of scale, are moving towards more formalization—others are reframing the vocabulary issue

22

Bottom-up control.

See Penn Tags <http://tags.library.upenn.edu/help/>

See the Library of Congress Flickr project <http://www.flickr.com/photos/library_of_congress/>

See The Art Museum Social Tagging Project <http://www.steve.museum/>

Libraries/Museums and Tagging

- Penn Tags
  - Still experimental, primarily internal to Penn
  - http://tags.library.upenn.edu/help/
- Library of Congress Flickr project
  - Open public tagging, still unclear how results will be used
  - http://www.flickr.com/photos/library_of_congress/
- The Art Museum Social Tagging Project
  - Research/software project focused on museum application
  - http://www.steve.museum/

23

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 for concepts

24

We already encode vocabularies for machine manipulation and other activities, for example the MARC authority and classification formats.

We already looked at MADS, another means of encoding the sorts of controlled vocabularies covered by the MARC authority format.

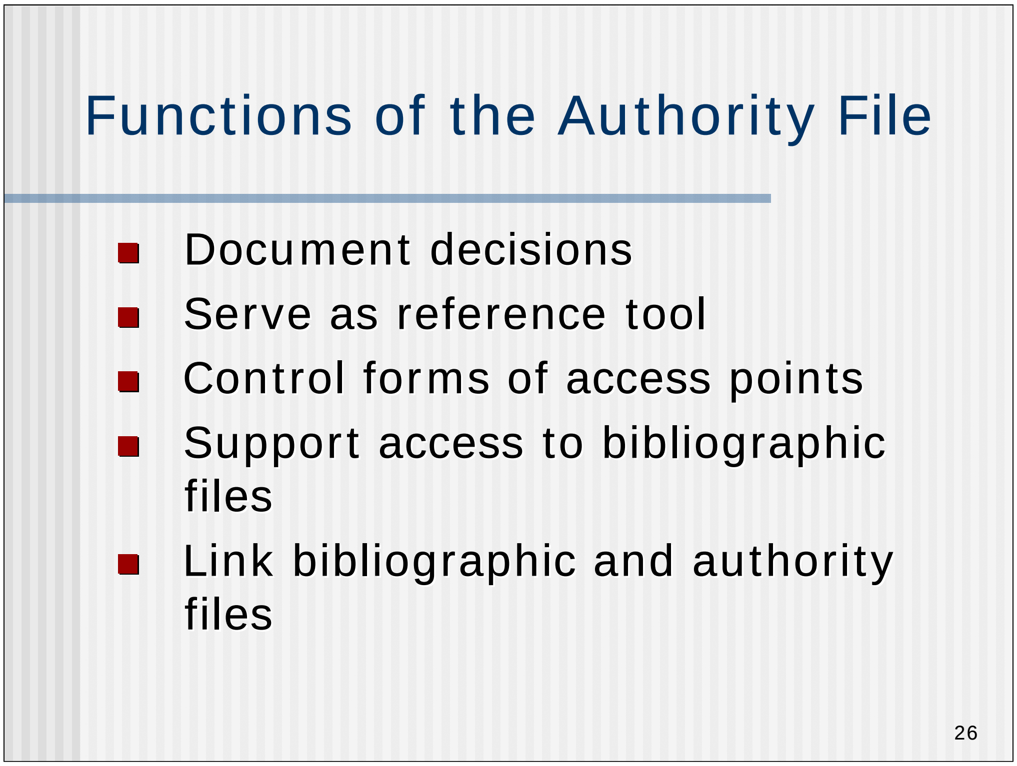
SKOS is used for topical kinds of concepts (we will learn more about this shortly).

New/Upcoming Standards:Authorities

- Functional Requirements for Authority Data (FRAD)
  - A new model for authority information
  - Developed by the IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR)
  - VIAF (Virtual International Authority File)
    - Prototype at: http://orlabs.oclc.org/viaf/
- A Review of the Feasibility of an International Authority Data Number (ISADN)
- Simple Knowledge Organization System (SKOS)—a W3C standard

25

FRAD is in process to be released in final form shortly; comments being added now from second review phase. The numbering of authority records is also a critical piece of the problem, and the paper outlines these issues.

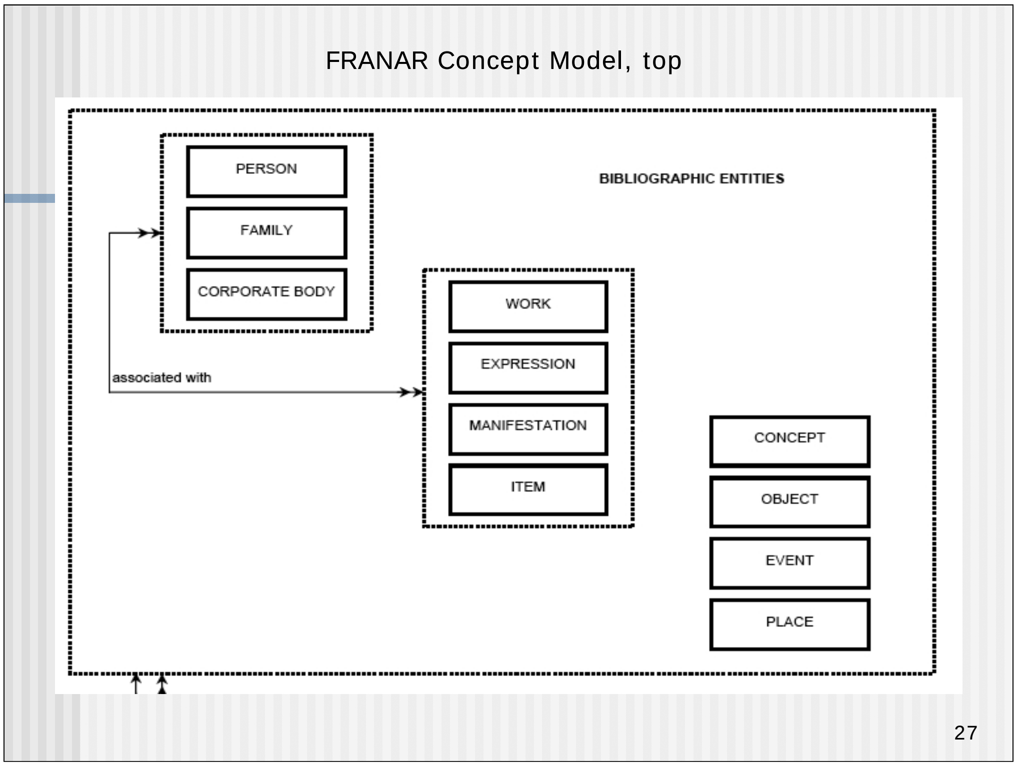


As a step toward understanding how authority data is used currently in the library context, the group has identified five functions of an authority file:

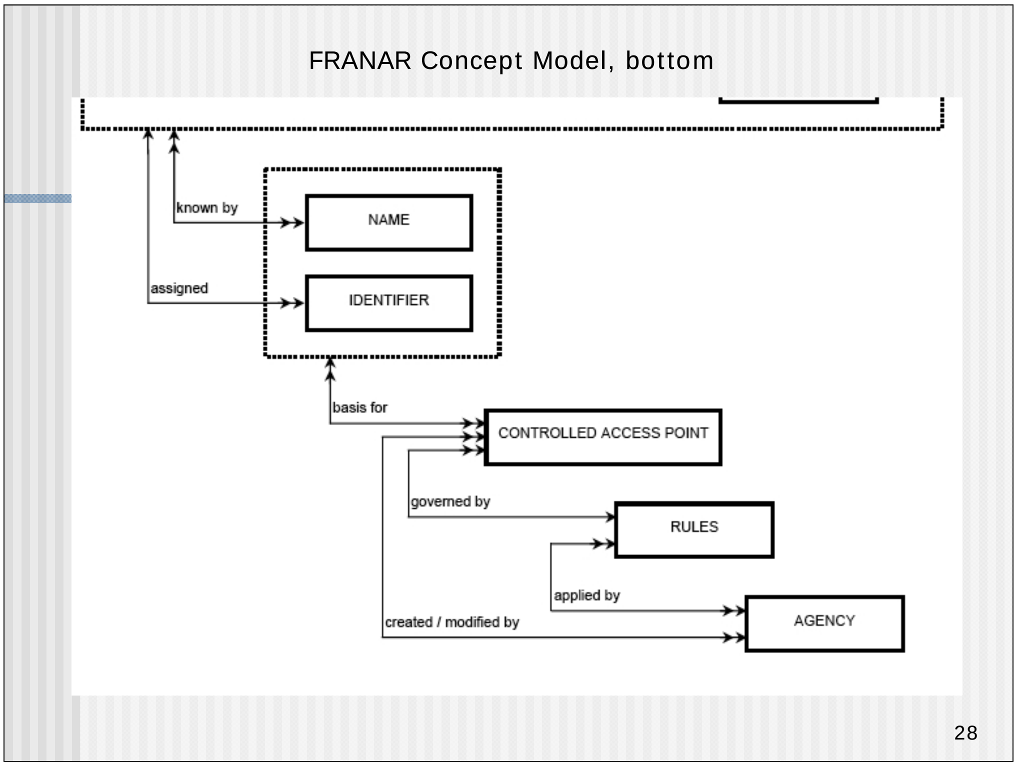
First, the authority file documents decisions made by the cataloger when choosing the appropriate controlled access points for a new bibliographic record or when formulating new access points.

Second, information in an authority file serves as a reference tool for those same two activities as well as providing information that can be used in distinguishing one person, corporate body or work from another. It may also serve to help the cataloger to determine that none of the access points in the authority file is appropriate and that a new access point is needed. It can also serve a broader reference function for other library staff.

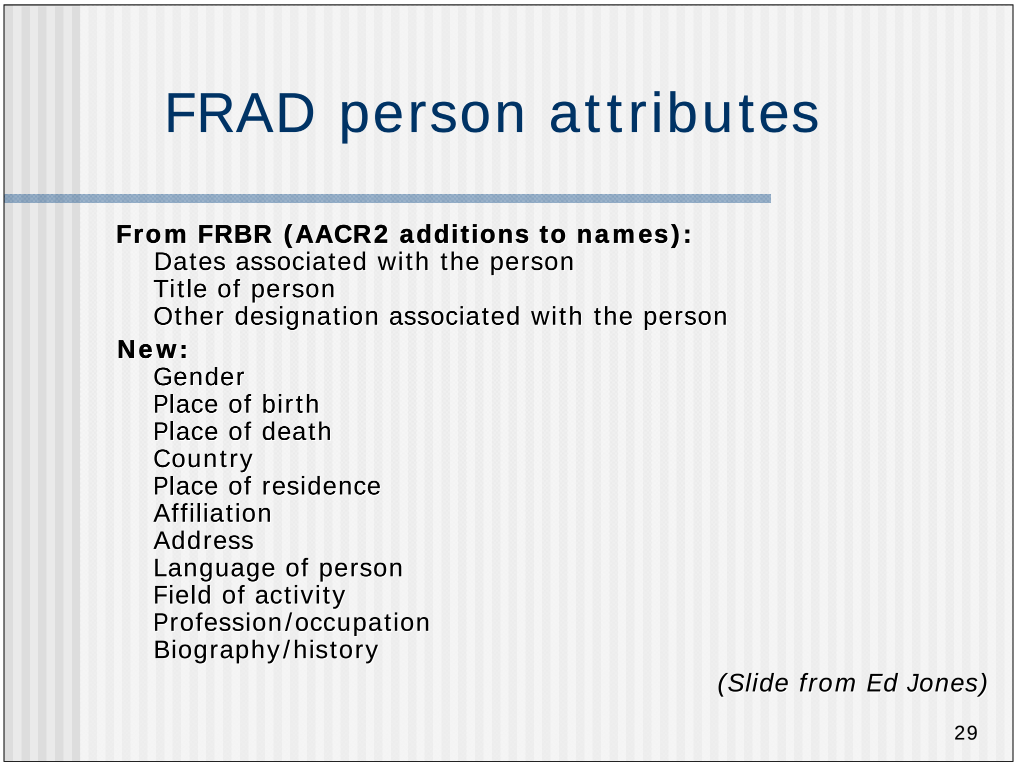
Third, the authority file can be used to control the forms of access points in bibliographic records and, in an automated environment, change those access points when the authority record itself is changed.



As is quite obvious, the FRBR model is the starting point for the FRANAR concept model—note on the bottom left hand corner that there are more relationships coming …



Here you can see that the name and identifier for the entities above are critical, no matter what the entity. These are related to the controlled access point, governed by rules, and created, modified and applied by an agency of some kind.

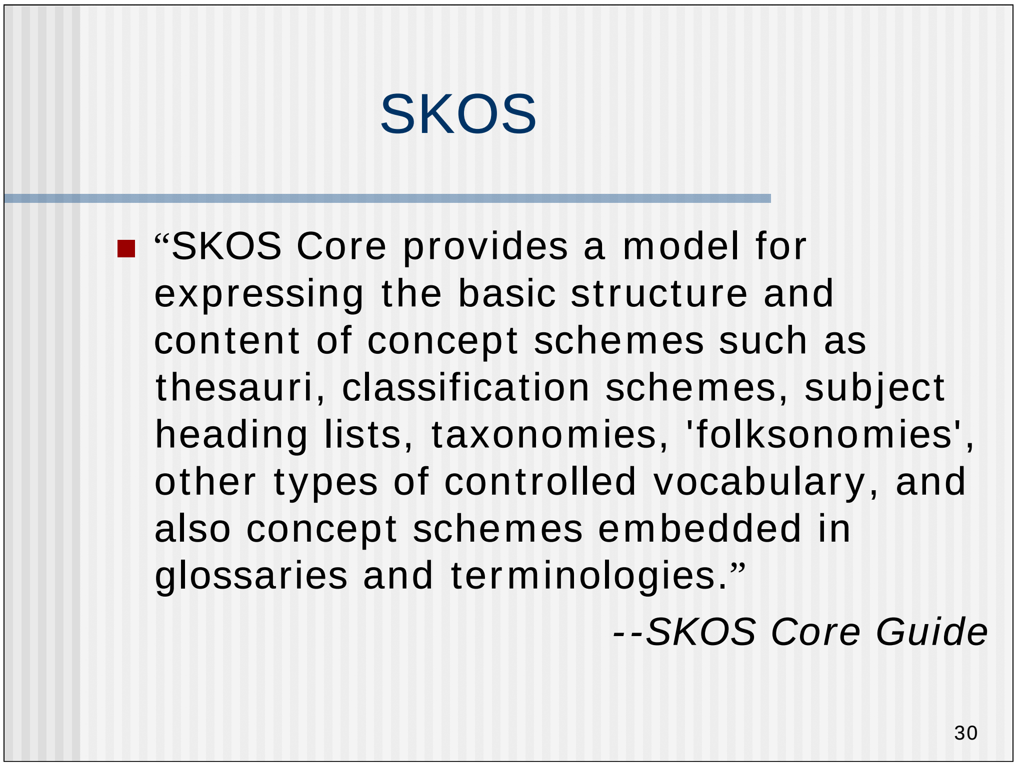


FRAD's got all the FRBR attributes—"additions to names" in AACR2 parlance—plus a passel of new ones

Where did these new attributes come from?

Gender ← UNIMARC Authorities

Others ← EU practice



Documentation for SKOS is in the bibliography. SKOS is intended to cover everything and more than what is included in the NISO thesaurus standard. SKOS is developed and maintained the W3C (World Wide Web Consortium) Semantic Web Deployment Working Group (SWDWG).

SKOS & RDF

- A World Wide Web Consortium (W3C) standard
- Based on RDF and OWL
- Data linked to and/or merged with other data
- Data sources distributed across the web
- http://www.w3.org/2004/02/skos/

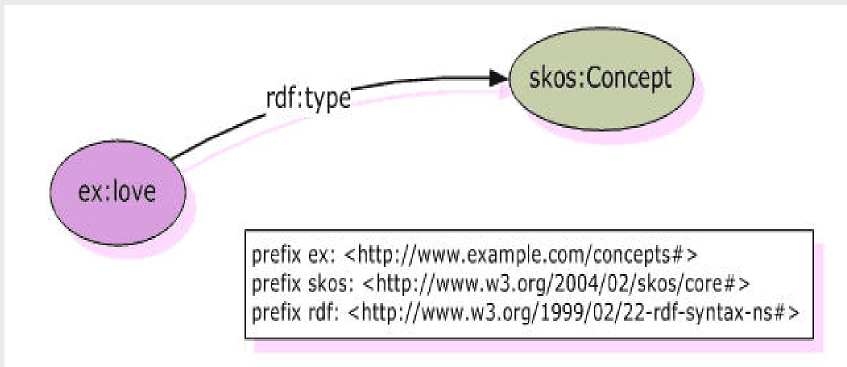
31

From SKOS Core Guide

As you can see, SKOS has a high level goal. It is still under development; most recent specification is *SKOS Reference* (2nd Working Draft, 9 June 2008).

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

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

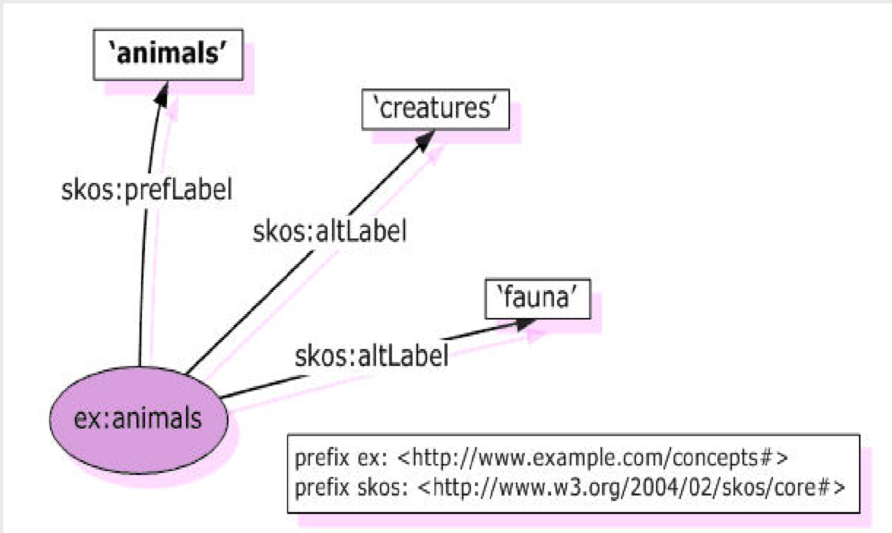


A SKOS conceptual resource is not a physical object.

"The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF), that can be used to express a concept scheme as an RDF graph. Using RDF allows data to be linked to and/or merged with other data, enabling data sources to be distributed across the web, but still be meaningfully composed and integrated."

--SKOS Core Guide

Preferred and Alternative Lexical Labels



"Animals" has a preferred label = animals

"Animals" has an alternate label = creatures

This could be used to generate text such as:

Creatures SEE Animals

Animals USE FOR Fauna

The RDF/XML Encoded Version

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <skos:Concept rdf:about="http://www.example.com/concepts#animals">
    <skos:prefLabel>animals</skos:prefLabel>
    <skos:altLabel>creatures</skos:altLabel>
    <skos:altLabel>fauna</skos:altLabel>
  </skos:Concept>

</rdf:RDF>
```

34

Concept is in this vocabulary

Preferred label is Animals

Alternate labels are Creatures, Fauna

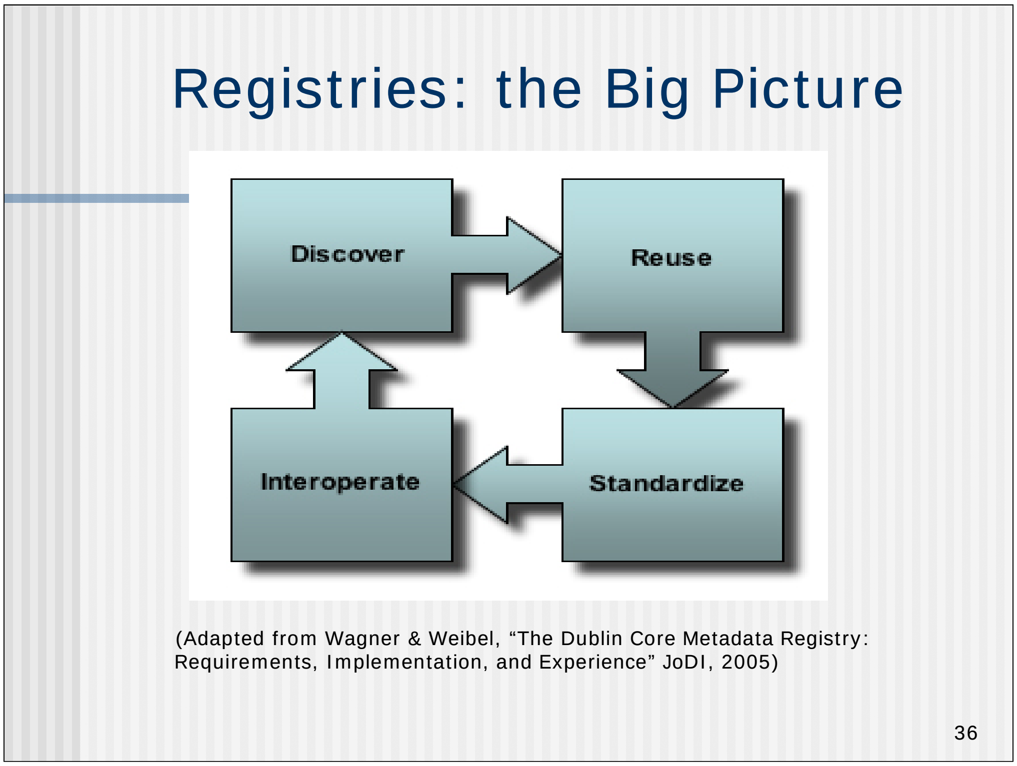
Example of ISO 639-2 language code in SKOS

```
<rdf:Description rdf:about=
"http://www.loc.gov/standards/registry/vocabulary/iso639-2/por">
<rdf:type  rdf:resource="http://www.w3.org/2008/05/skos #Concept"/>
        <skos:prefLabel xml:lang="x-notation">por</skos:prefLabel>
        <skos:altLabel xml:lang="en-latn">Portuguese</skos:altLabel>
        <skos:altLabel xml:lang="fr-latn">portugais</skos:altLabel>
        <skos:notation rdf:datatype="xs:string">por</skos:notation>
        <skos:definition xml:lang="en-latn">This Concept has not yet been
        defined.</skos:definition>
        <skos:inScheme rdf:resource=
        "http://www.loc.gov/standards/registry/vocabulary/iso639-2"/>
        <vs:term_status>stable</vs:term_status>
        <skos:historyNote rdf:datatype="xs:dateTime">2006-07-
        19T08:41:54.000-05:00</skos:historyNote>
        <skos:exactMatch rdf:resource=
        "http://www.loc.gov/standards/registry/vocabulary/iso 639-1/pt"/>
        <skos:changeNote rdf:datatype="xs:dateTime">2008-07-
        09T13:49:05.321-04:00</skos:changeNote>
 </rdf:Description>
```

This is one type of SKOS expression, which is RDF/XML. Tags defined by SKOS a wrapped in an RDF wrapper. skos:prefLabel uses the language code as value so as to not give a preference to a term in a particular language. The skos:altLabel is used here for the various language names

skos:inScheme tells you what concept scheme this entry is included in; there can be multiples.

skos:exactMatch gives a URI for the other code which is exactly the same; in this case there is the 2-character code “pt” (ISO 639-1), which is equivalent to this one.



How does all of this encoding of these concepts, etc. work?

Coupled with registries, it allows you to do certain things.

We've been able to support traditional library vocabularies because of institutional commitments. But for others, use registries.

Between Interoperate & Discover: Express

Between Standardize and Interoperate: Register

Could use MADS or SKOS to expose vocabularies. SKOS is better for subjects and handles concepts well.

E-Card is an encoding standard for names

If you want vocabularies to be available for use by others, you must document and publish them.

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 different metadata schemes

37

You can (should) register metadata schemes, values used in your controlled vocabularies.

Registries may also be used to support crosswalks and translations.

Registries may assist in coping with different metadata schemes if machine readable crosswalks are available.

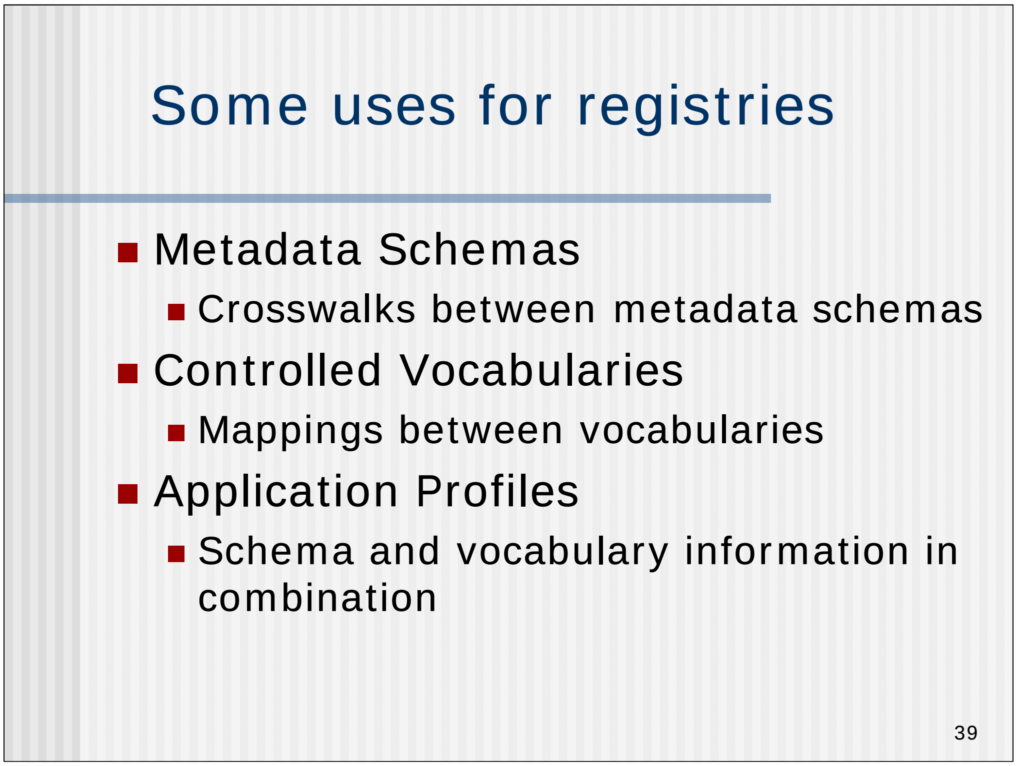
Declaration, documentation, publication

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

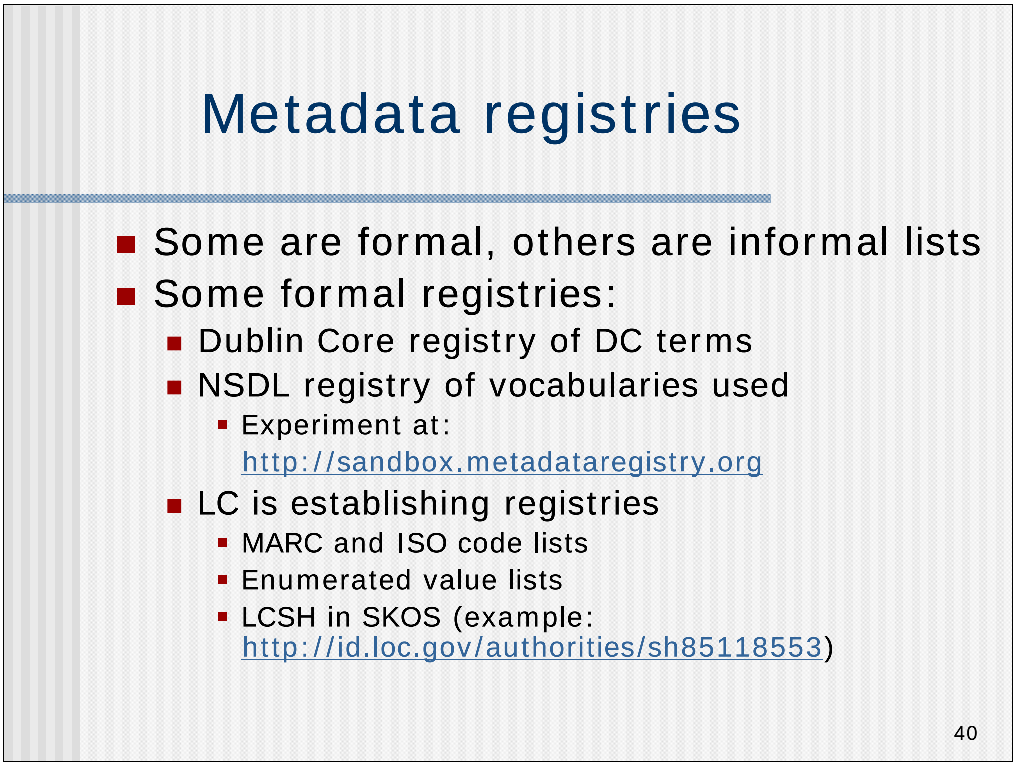
38

These are the pieces that are important for building the infrastructure that makes good metadata function.

Most of these steps can be done by anyone. Some are building up our support this way– distributed support, rather than centralized support. Even the registry portions (still being built) will be distributed. There is also a place for centrally published controlled vocabularies, where a maintenance agency takes control for a given standard.



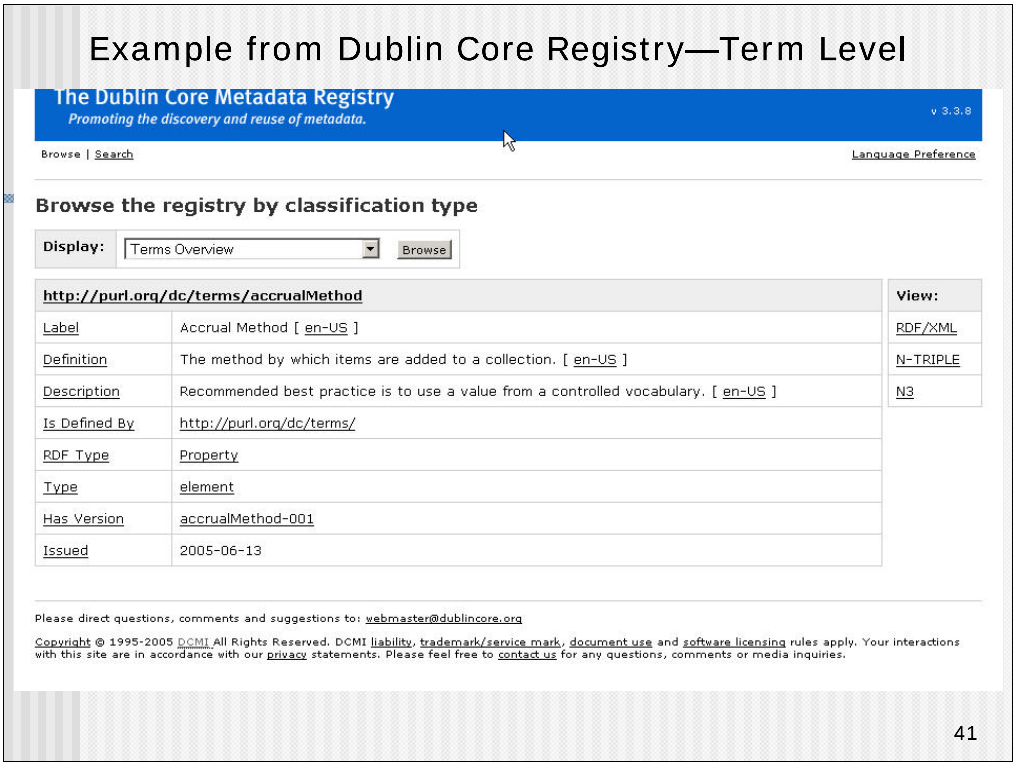
Although many registries are now only registering schemas, the intention is that all these pieces need registration and management for the benefit of all.



We will be seeing more work in the future in this area. Formal registries allow reuse of data standards in a broader community than that which developed it.

Click on the link for LCSH in SKOS to see a prototype of how LCSH in SKOS might be used.

Remind the trainees that there is a whole other course on controlled vocabularies as part of the Cataloging for the 21st Century series of workshops.



Gives terms, definitions, translations into other languages, versions. Optimized for multi-language capability.

Can view in different encodings (RDF/XML for example).

Dublin Core manages the DCMI type vocabulary—this registry is intended only for DC terms currently.

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

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

Metadata Standards and Applications

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

Efficiencies are achieved by decades’ worth of developing agreements and shared models.

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

“Submission” model developed with the idea that people who are submitting information will create the metadata.

Assumption that the submitter can reliably create their own metadata.

Usually done at universities with professors as authors.

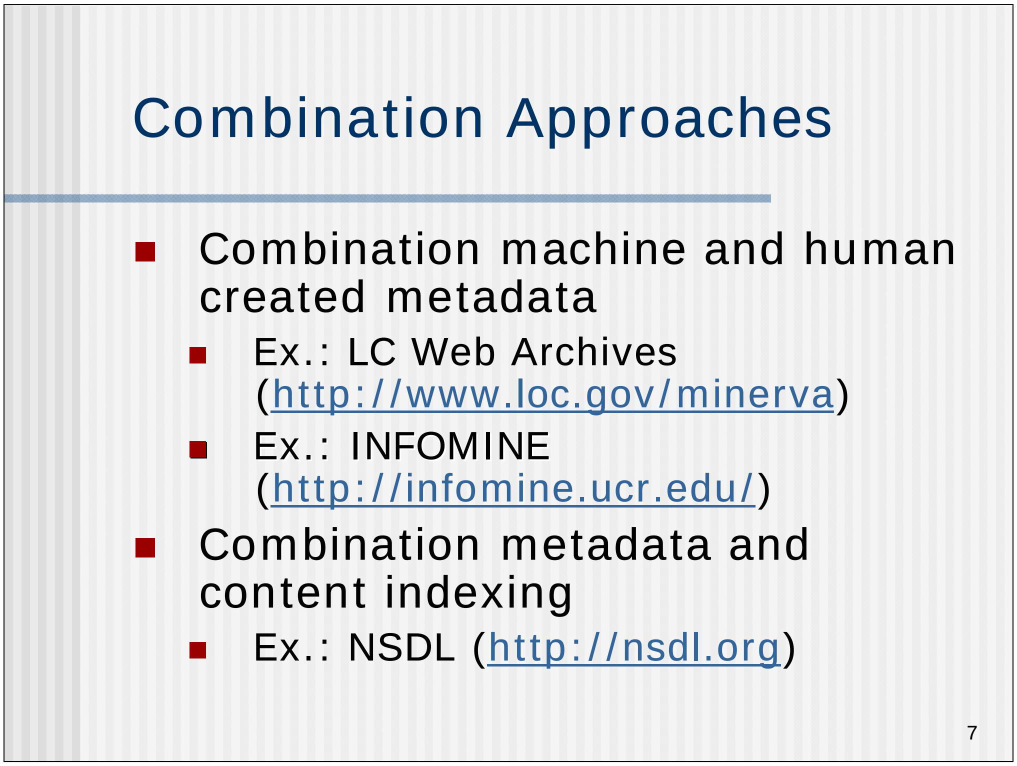
The Automated Model

- Based largely on text analysis; doesn't usually extend well to non-text or low-text
- Requires development of appropriate evaluation and editing processes to support even minimal quality standards
- Still largely research; few large, successful production examples
- One simple automated tool to try: http://www.ukoln.ac.uk/metadata/dcdot/
- Automated model may be more efficient for for technical metadata

6

Some use natural language processing, text analysis. At least, machines are consistent. It's hard to use a purely automated approach.

Also works for technical metadata.



The machine and human approach is used at LC for the Web Archives collections.

MINERVA: For the various collections, tools were used to extract some metadata. Rights information was extracted from the Leaderboard (the project's permissions tracking tool). Other metadata was extracted from the Web "ARC" files (a file format used for harvested websites used by the Internet Archive). Additional metadata was boiler-plated in when it applied to all records (e.g. the genre "website").

Some descriptive and technical metadata can be extracted from the file, e.g., file type…

Preliminary keyword, title, and subject metadata were extracted from the archived Web sites to create preliminary MODS records that were subsequently reviewed and/or enhanced by catalogers who assigned controlled subjects from Library of Congress Subject Headings (LCSH) or Thesaurus of Graphic Materials (TGM). A Lucene search interface was developed to search the MODS records both within and across the archived collections.

NSDL—link from metadata provides access to indexing software; index data used in combination with metadata to provide access

In practice, it has not been honed well. Users get unanticipated results but don't know why and are sometimes frustrated by this.

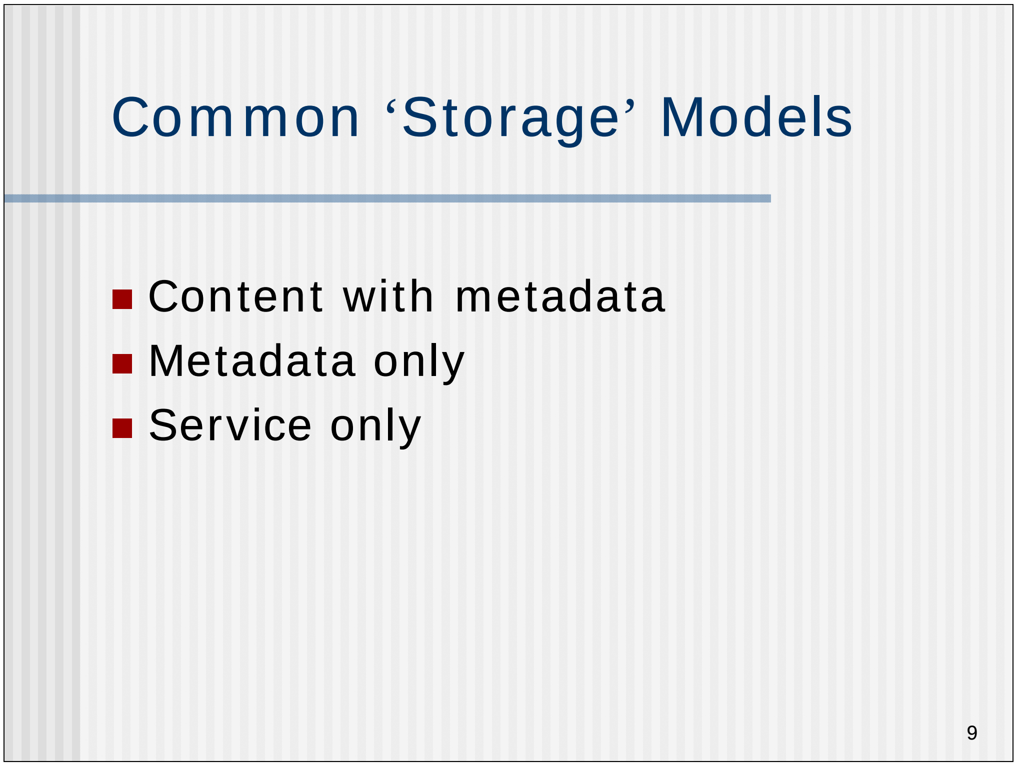
Content “Storage” Models

- ‘Storage models’ 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

We come from an environment in which content and metadata are separate things. In the digital world, that’s not always the assumption. Decisions about that relationship affect access, etc.

We have seen how metadata can be packaged w/ the objects in digital library applications and the way metadata can be used in retrieval and management.



Content management systems: metadata stored with content

A METS file bundles together content with metadata (by either embedding or providing links to files)

Metadata only example: library catalogs

Will talk about "service only" model as we go along

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),), a full-text markup language (as an example of an application, see the Comic Book Markup Language at http://www.cbml.org/)
- Often proves difficult to update and not optimized to manage metadata over time

10

There are not many HTML “meta” tags embedded. Search engines don’t’ use them because spammers and others have abused them.

Metadata sometimes is not flexible enough—doesn’t allow repeatable elements, often optimized to managing the content and not the metadata. Bibliographic information is not static information.

Image files have technical metadata w/content. e.g., size, format, settings…TIF, GIF, JPG

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

There might be links to services like licensed content.

Traditional cataloging in library catalogs is usually title level, not at the article level.

Metadata aggregations are analogous to OCLC or RLG.

API: Abbreviation of ***a**pplication **p**rogram **i**nterface*, a set of routines, protocols, and tools for building software applications (definition from Webopedia)

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)

This metadata may be reused, converted to new formats, and packaged with content (e.g., METS)

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

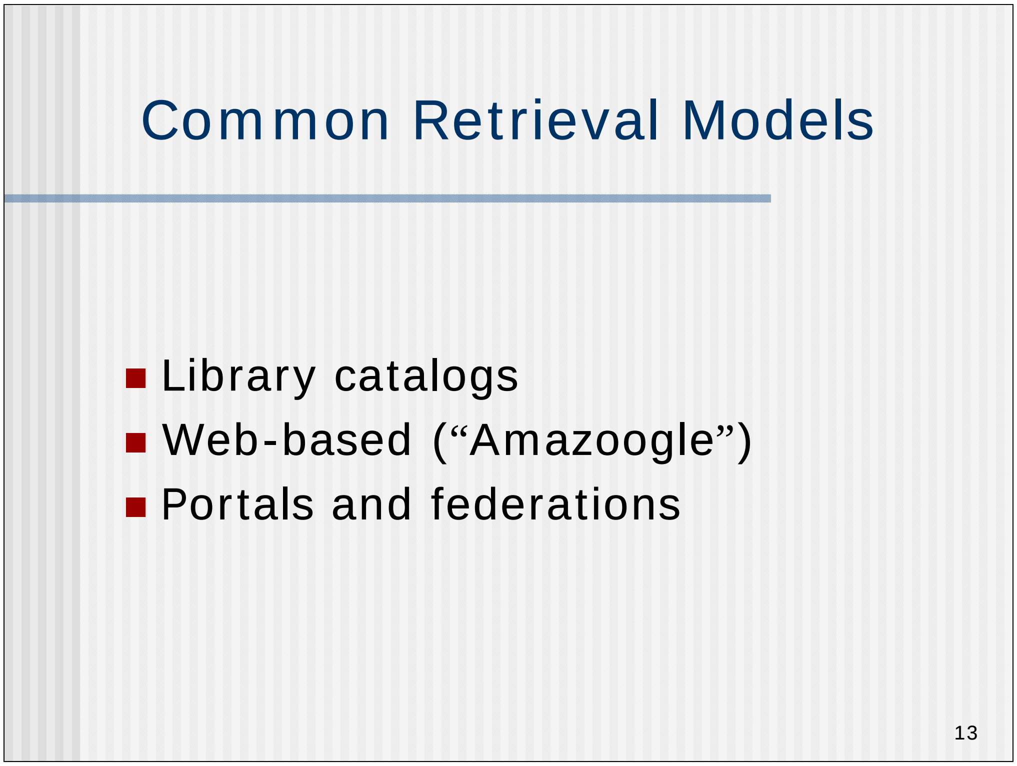
Think about: The services mentioned here are supported partially or fully by metadata.

Google: think of “page rank” as a form of metadata (not descriptive as we’re used to). Yahoo began as a more “library-like” service with human categorization but now uses more of a search model.

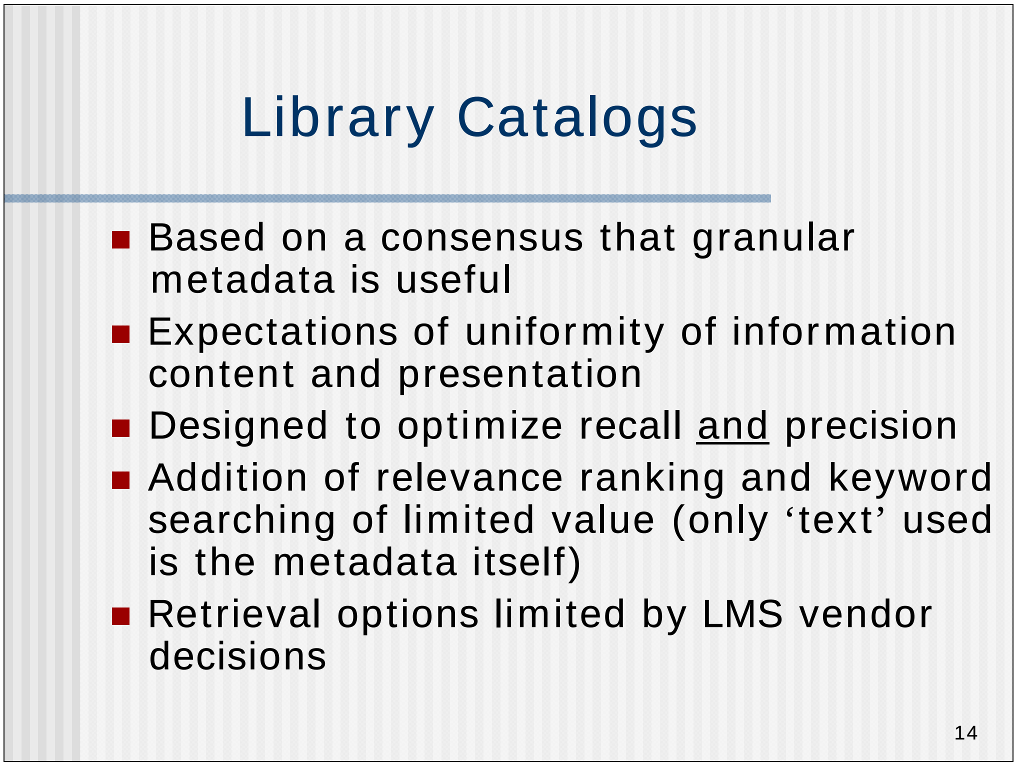
“Website powered by Google” an example of distributed service software.

Will talk a little later on about how link resolvers are often metadata based

Non-distributed metadata can’t be incorporated directly into your own services



“Amazoogle” = Amazon and Google



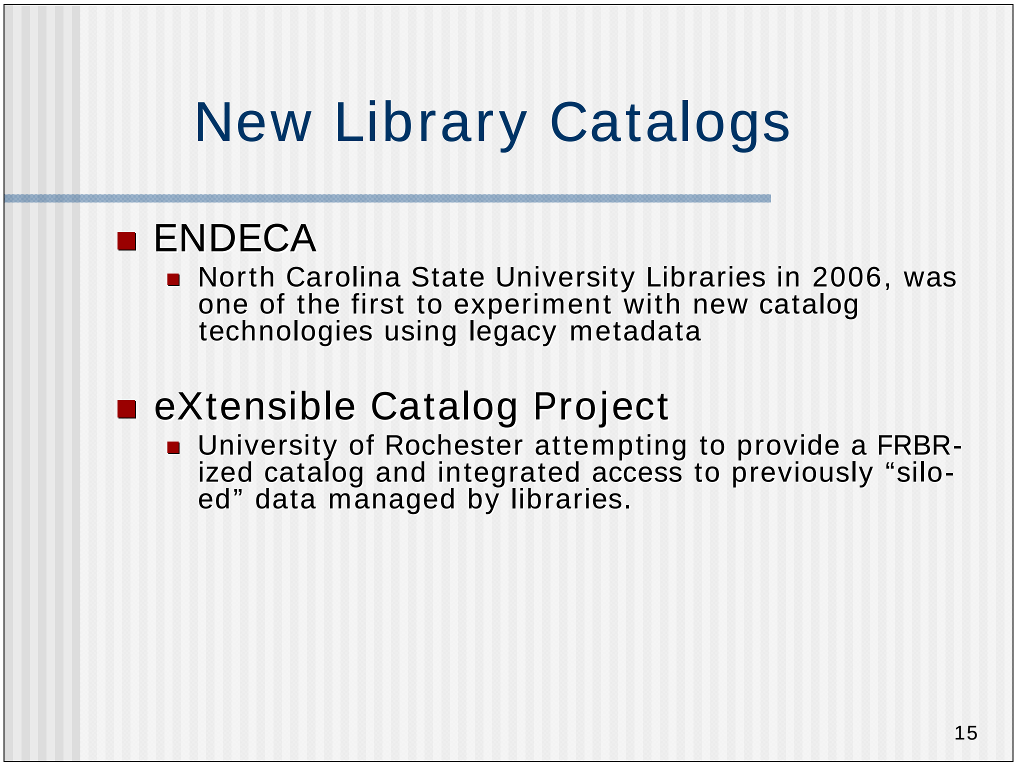
We catalog things at a particular level of granularity and not much above or below that (books, not chapters; serials, not articles). This can be confusing to users who might expect more integration with other search services.

Movement towards unbundling library management systems and separating the user interface from the back-end management is in early stages but worth watching. OCLC's WorldCat Local is being explored by some libraries as an alternative to the OPAC. There are other efforts trying to enhance the user experience.

We have traditionally cataloged things at a particular level of granularity and not much above or below that. We're finding ways to link to records which are more granular. Things are evolving…

Keyword searching is limited to the text of metadata, not the text of the object, which limits its usefulness, particularly for those use to a Google approach.

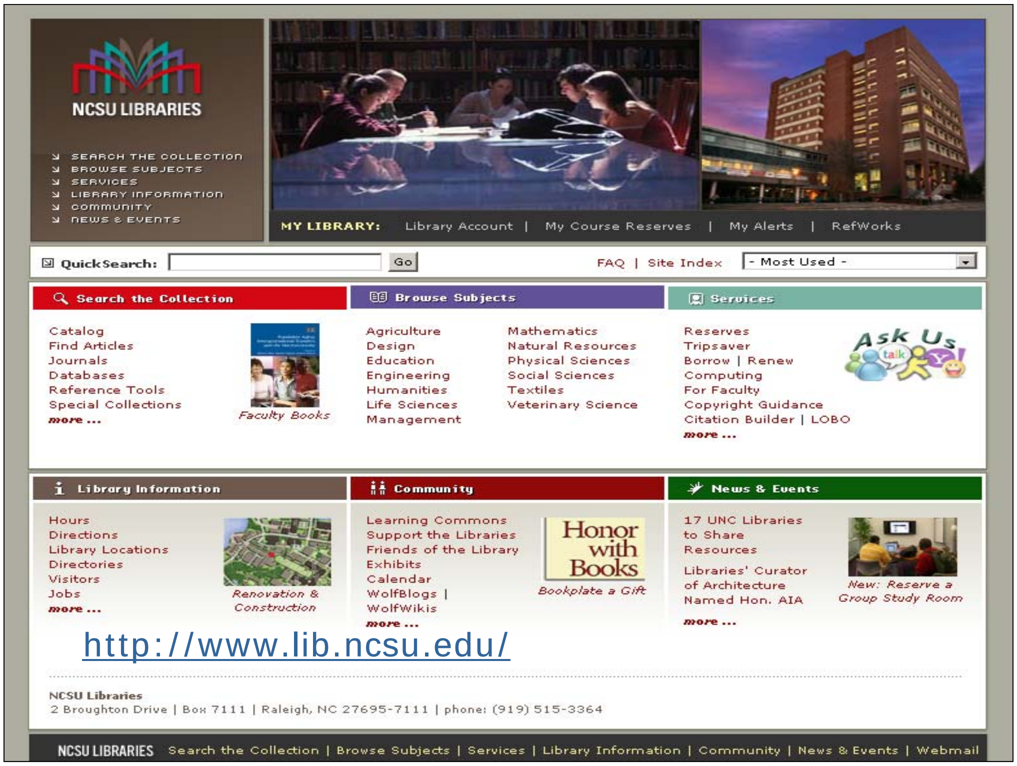
Movement towards unbundling library management systems and making catalogs better is in early stages but worth watching.



Endeca raised $10M that it used to operate as a research company looking at new ideas and opportunities for overcoming the constraints of relational database technology. Their goal was to build a software platform capable of optimizing and summarizing huge volumes of data that could ultimately be used by "everyday people for exploration," without requiring armies of experts.

The Endeca Information Access Platform is powered by a new class of database, **the MDEX Engine**™, that helps you successfully build tailored applications for people to explore your existing data, regardless of its source or format.

In early 2006, the NCSU Libraries announced the first library deployment of a revolutionary new online catalog. Leveraging the advanced search and Guided Navigation® capabilities of the Endeca ProFind™ platform, the NCSU Libraries' new catalog provides the speed and flexibility of popular online search engines while capitalizing on existing catalog records.



Students, faculty, and researchers can now search and browse the NCSU Libraries' collection as quickly and easily as searching and browsing the Web, while taking advantage of rich content and cutting-edge capabilities that no Web search engine can match.

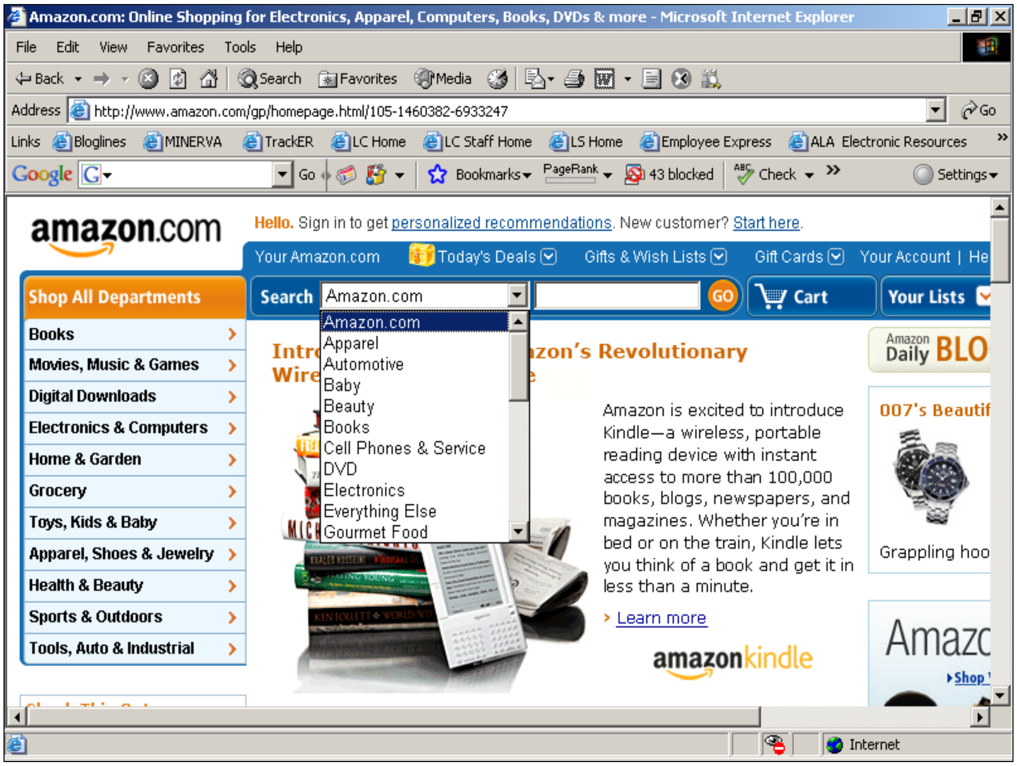
Go live and search through the system...

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

Amazon is a combination approach. Some metadata is actually ONIX, but they are also combining that basic metadata with other “stuff” including full text and user-supplied information.



Amazon page example show the "Google type" search box connecting the various categories, and browse features. Meta tags are used to provide this searching.

Most people are familiar with Amazon, so no real need to go live.

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

Sometimes users can search for free but can't get to the content in protected metadata without a license or registration. Some protected services had the notion that there would be an income stream from their metadata.

Portals to the World: A Library of Congress collection of electronic resources from around the world. A Collection level MARC record for each country is provided in the LC OPAC.

The Portal can be accessed by an engine search on "Portals to the World."

BEN requires registration to search or use resources, and many of the resources have specific rights attached. They do not distribute their metadata to other partners, nor is it available to search engines.

XML based digital library application

- Similar to a portal application
- May use a database for record creation and maintenance
- Often uses open source tools
- Files are indexed for searching and presented on the Web using an XML based publishing framework
- Combines some of the other metadata creation, storage and retrieval approaches
- http://www.loc.gov/performingarts/

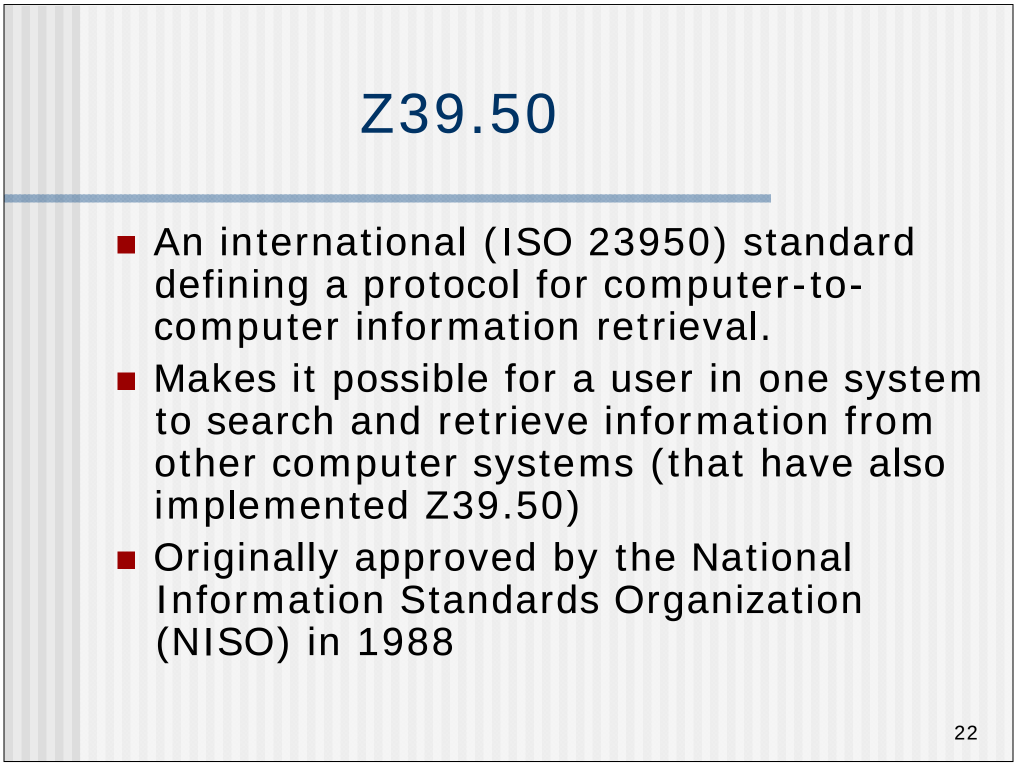
20

This seems to be where digital collection development is heading.

Last bullet links to an example of this sort of application, "**Performing Arts Encyclopedia : Explore music, theater, and dance at the Library of Congress."** Visit and study the site if you decide to go live or feel free to use a different example.

There are standards that help manipulate and navigate through the various metadata standards. A few worth mentioning are:

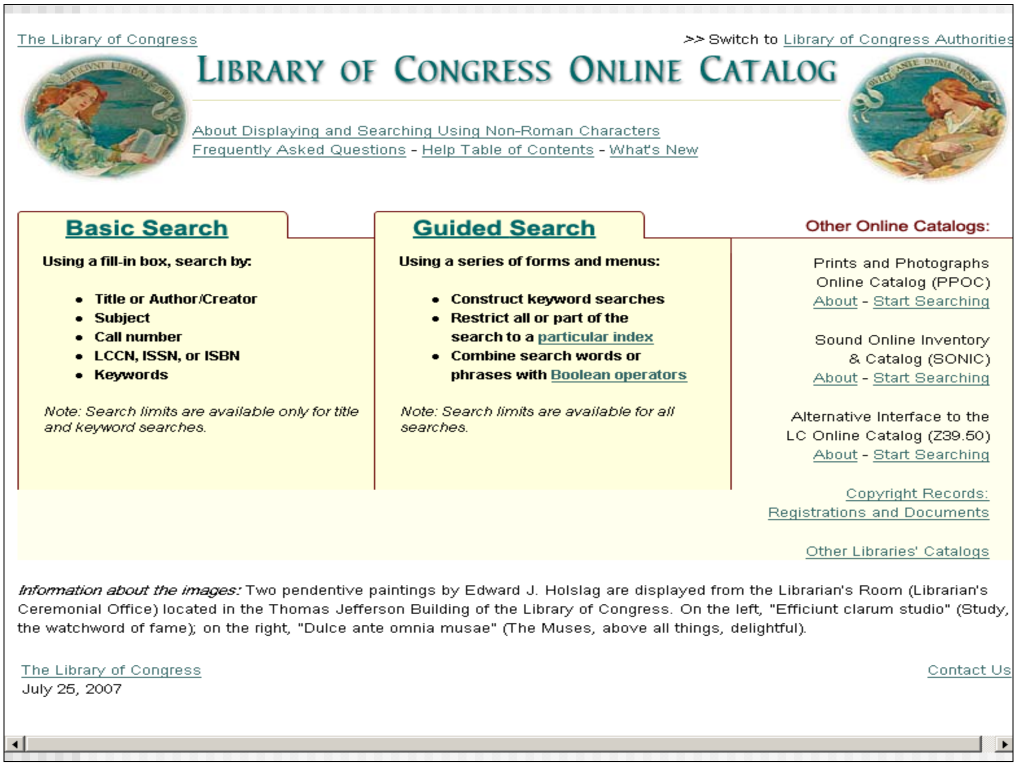
Z39.50, SRU and metasearch.



Z39.50 is a national and international standard defining a protocol for computer-to-computer information retrieval. Z39.50 makes it possible for a user in one system to search and retrieve information from other computer systems (that have also implemented Z39.50) without knowing the search syntax that is used by those other systems.

As a network application standard, Z39.50 is an *open* standard that enables communication between systems that run on different hardware and use different software. The Z39.50 standard was developed to overcome the problems associated with multiple database searching such as having to know the unique menus, command language, and search procedures of each system accessed. Z39.50 simplifies the search process by making it possible for a searcher to use the familiar user interface of the local system to search both the local library catalogue as well as any remote database system that support the standard.

Z39.50 has been implemented throughout the US and Europe by libraries, library software vendors, and bibliographic utilities such as OCLC. It is being succeeded by SRU (Search and Retrieve via URL).



For example, from the LC OPAC you have the option of searching just LC's catalog,

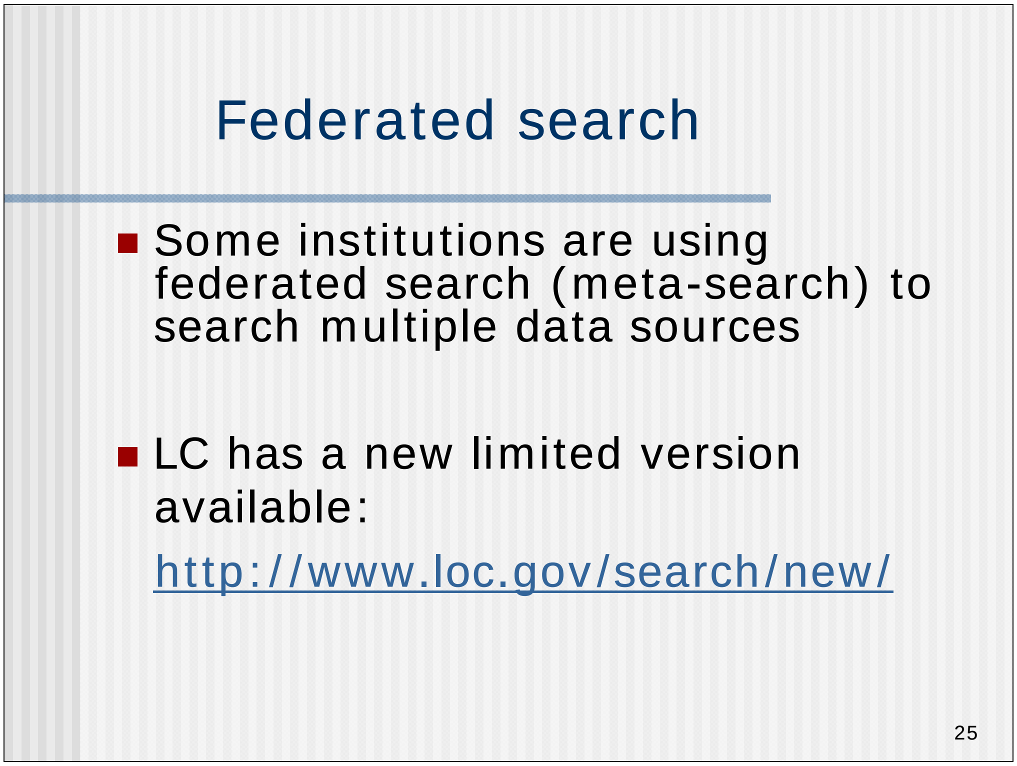
And on the right-hand side of the page you can see there is a listing of "other online catalogs." If you are interested in searching the databases of various institutions in just one query, you could go to Z39.50 gateway (third one down).

SRU is intended to define a standard form for Internet search queries as well as the structure of the responses.

SRU is the Web Services-based protocol for querying Internet indexes or databases and returning search results.

As metadata specialists, you may hear of these standards as you work with the technical specialists on digital projects. You don’t need to know it, but being familiar with the technical jargon is useful.

SRU is being used as a search and retrieve protocol that is XML based, so can output results in any metadata format that uses XML.



Federated or Meta- search (also called *integrated search*) is the ability to simultaneously search multiple data sources -- typically including the Internet and corporate intranets and databases as well as hard drives and removable storage on the user's computer -- from a search term entered into a text box on the desktop.

Metasearch services rely on a variety of approaches to search and retrieval including open standards (such as NISO's Z39.50), proprietary API's, and screen scraping. However, the absence of widely supported standards, best practices, and tools makes the metasearch environment less efficient for the system provider, the content provider, and ultimately the end-user.

LC has a version of a meta-search available on in Web site. All are invited to visit, experiment and provide comments.

[Instructors should visit: http://www.loc.gov/search/new/help/faq.html#faq1 or feel free to use a different example]

This can be done as a class exercise led by the instructor.

See extensive notes with the Instructors' manual.

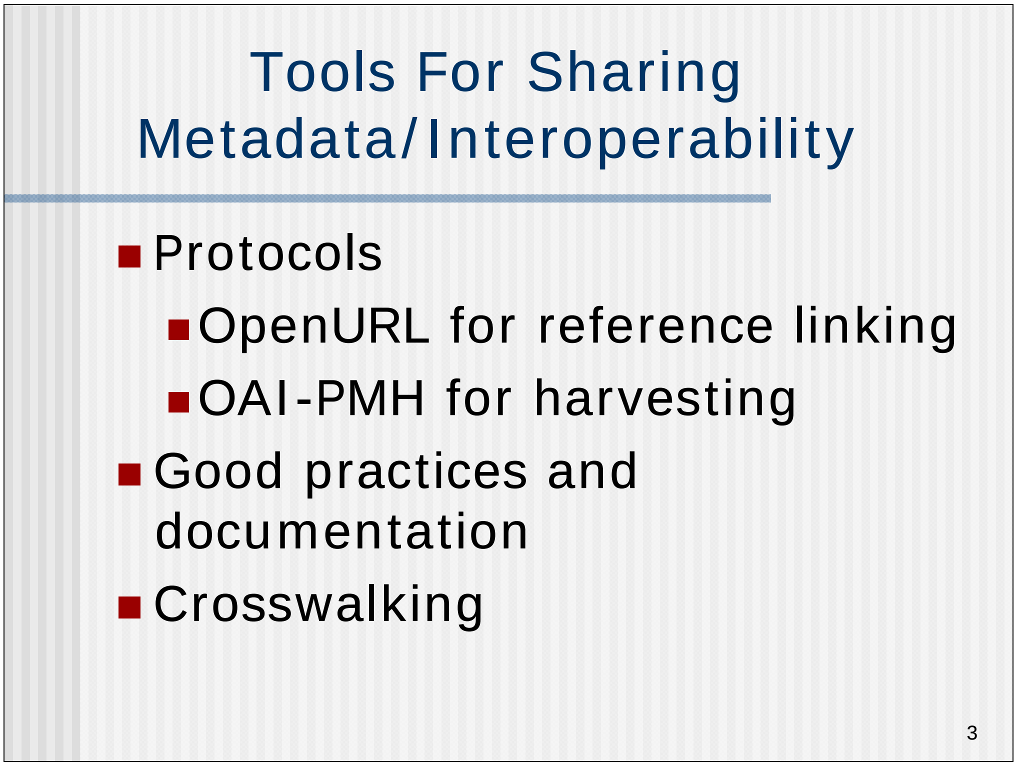
8. Metadata Interoperability and Quality Issues

Metadata Standards and Applications Workshop

Goals of Session

- Understand interoperability protocols (OpenURL for reference, OAI-PMH)
- Understand crosswalking and mapping as it relates to interoperability
- Investigate issues concerning metadata quality

2



Interoperability

> the ability of two or more systems or components to exchange information and to use the information that has been exchanged

A **repository** is a collection of resources that can be accessed to retrieve information.

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

Interoperability is a kind of buzzword but stands for something that traditional libraries have taken for granted. The metadata world is still exploring what it means.

Users are trying to figure out certain things (see the questions on the slide)

The slide asks important questions for resource creators.

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)

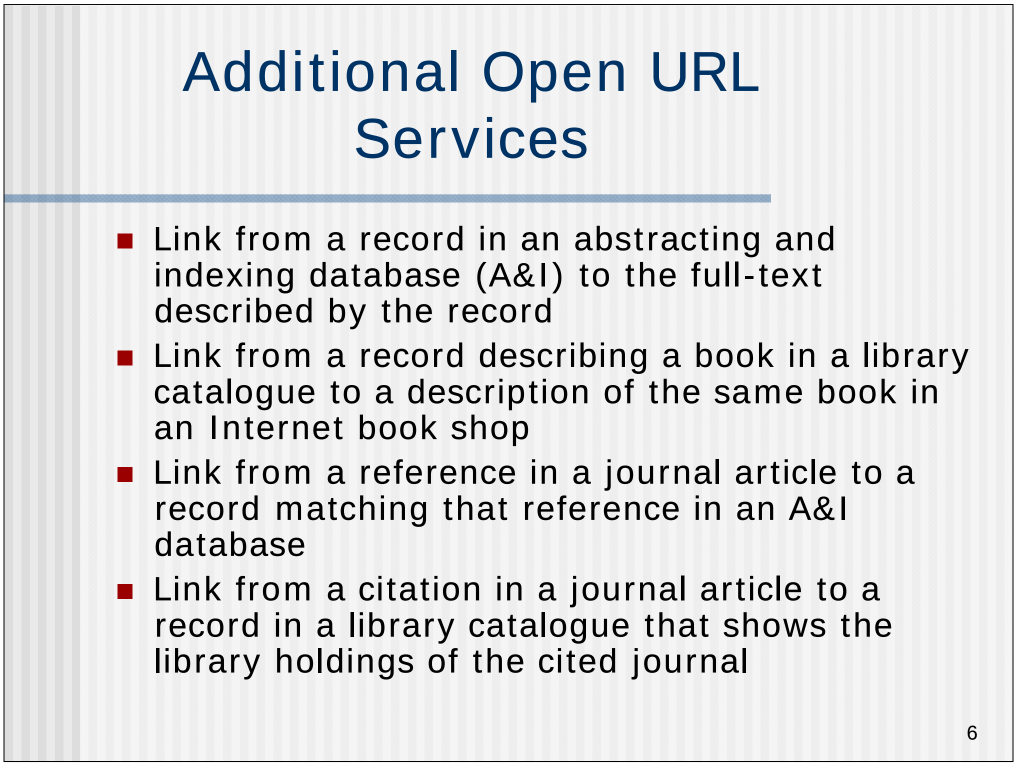
5

The best way to understand OpenURL is to look at one and take it apart. We’ll do that in an OpenURL demo in a few minutes. The initial intent behind OpenURLs was to use a ubiquitous protocol like “http” to pass journal-citation information around. It’s interesting to note that both OAI and OpenURL were developed in part by Herbert von de Sompel and both use “http” as a basis for conveying data.

[Ask the participants: How many of you are from institutions that use OpenURL? How many of you have worked to get the protocol to work? How many of you generally understand what OpenURL does? Use that as a basis for how deeply you need to cover OpenURLs.]

Immediate use of OpenURL: to go from a citation for a journal article to the full text for that article that is available and licensed for your users.

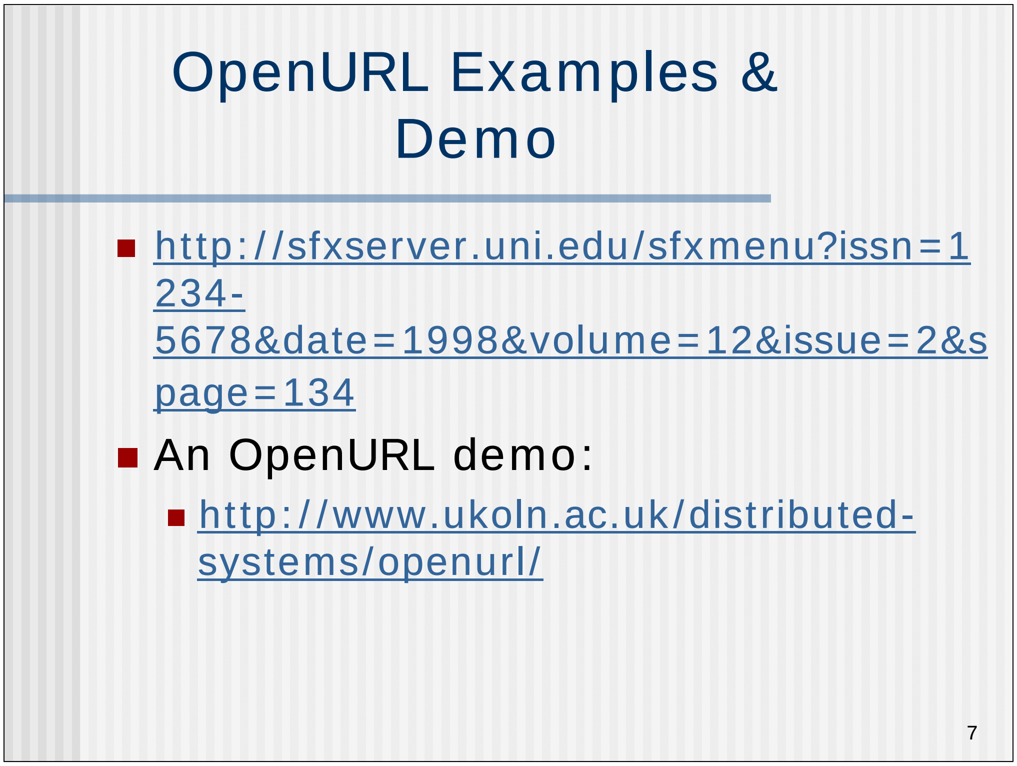
The “most appropriate copy” can be set locally, based on resources that you have access to . For example, if you have access to a journal from Elsevier, SicenceDirect, and Emerald, you might choose to give the user the Elsevier “version” first.]



Examples:

You can use citation information to see if the resource is available in print at the local library. Or you can populate an ILL form.

Or you can send the citation information to a commercial vendor.



The top bullet is an example of an “OpenURL” url. It includes: ISSN, Date, Volume number, Issue number, Start page number. The full text services vary for what you need. [At the time of this revision, the URL did not work, but this is intended to show what an openURL looks like.]

Click on the demo URL:

If the examples on the right side of the page are not activated, click on OpenResolver (demo) on the left side of the page. Typically, there will be some kind of button that is turned on for local access. Usually, if there is a problem with resolving a link, it has to do with the source metadata that is transmitted via the OpenURL protocol rather than with the protocol itself. IDs like URIs are important.

Choose the first article from the list on the right side of the page. In this example, there is no start page because D-LIB is not a print publication. In an “advanced” implementation of OpenURL, you can see services like ILL, local catalog lookup, download into a citation management software.

If someone does a search in Google Scholar, based on an IP address, it will turn on your link resolver in Google Scholar to find the item in your local catalog.

This version of OpenURL focused almost exclusively on the article resolver. The current standard is more open and can be used for broader applications. You can represent URLs in XML. You can pass multiple citations in packages. You can define your own protocols for specific users and specific materials. For example, classical texts like the Illiad, Odyssey, or the Bible could be resolved to the sentence or paragraph level using the standard numbering in the texts themselves.

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”

8

Interoperability is of particular issue when we talk about distribution outside the source of the metadata. One important resource for interoperability is OAI-PMH

The effort began with preprints in computer science and physics but the applicability is broader than that.

Metadata is created not just to convey bibliographic information (for example, author information) but to serve a number of functions within digital libraries.

The open archives approach is to enable access to Web-accessible material through interoperable repositories for metadata sharing, publishing and archiving.

OAI-PMH in a Nutshell

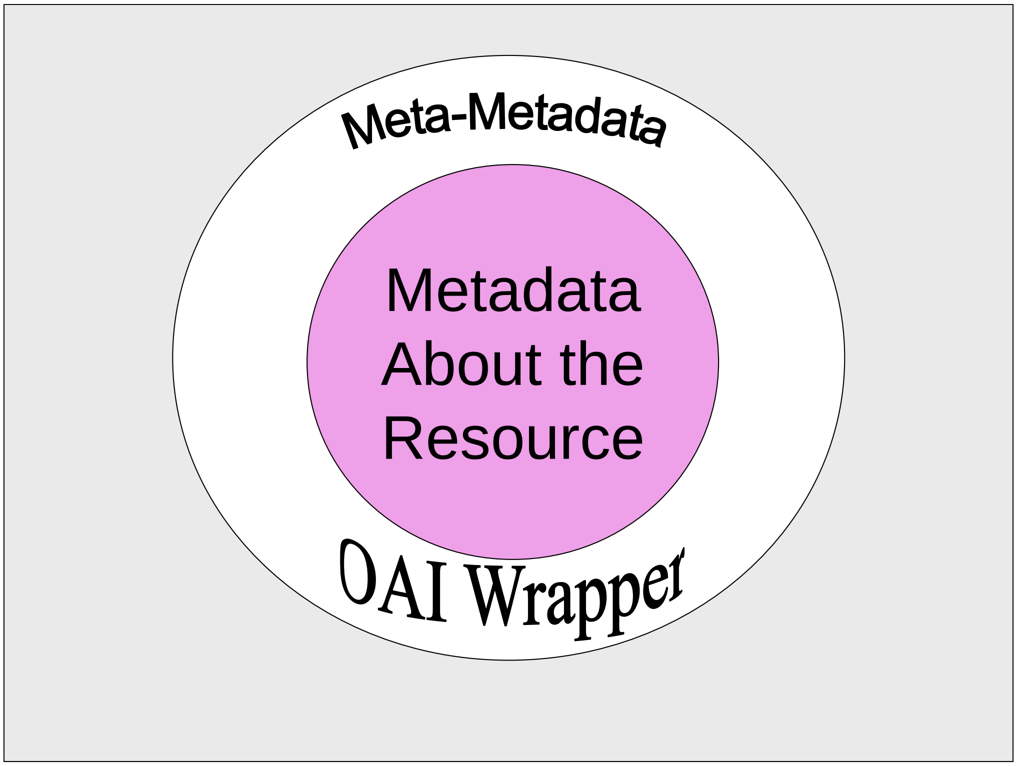
- Essentially provides a simple protocol for “harvest” and “exposure” of metadata records
- Specifies a simple “wrapper” around metadata records, providing metadata about the record itself
- OAI-PMH is about the metadata, not about the resources

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

9

Protocol for interaction between a harvester and information exposed for harvesting.

Provides administrative metadata about the resource (DC has no administrative metadata) such as ID for the metadata record.



A graphic view of the relationship of OAI with the descriptive metadata about the resource.

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)

11

OAI_PMH is a mechanism for distributing metadata.

Technical support varies between institutions, so barriers to participation are a big issue. Important to note that servers do not need to be dedicated to OAI support, and many software packages are available via the OAI website. A predictable flow of data maximizes the machine flow of information.

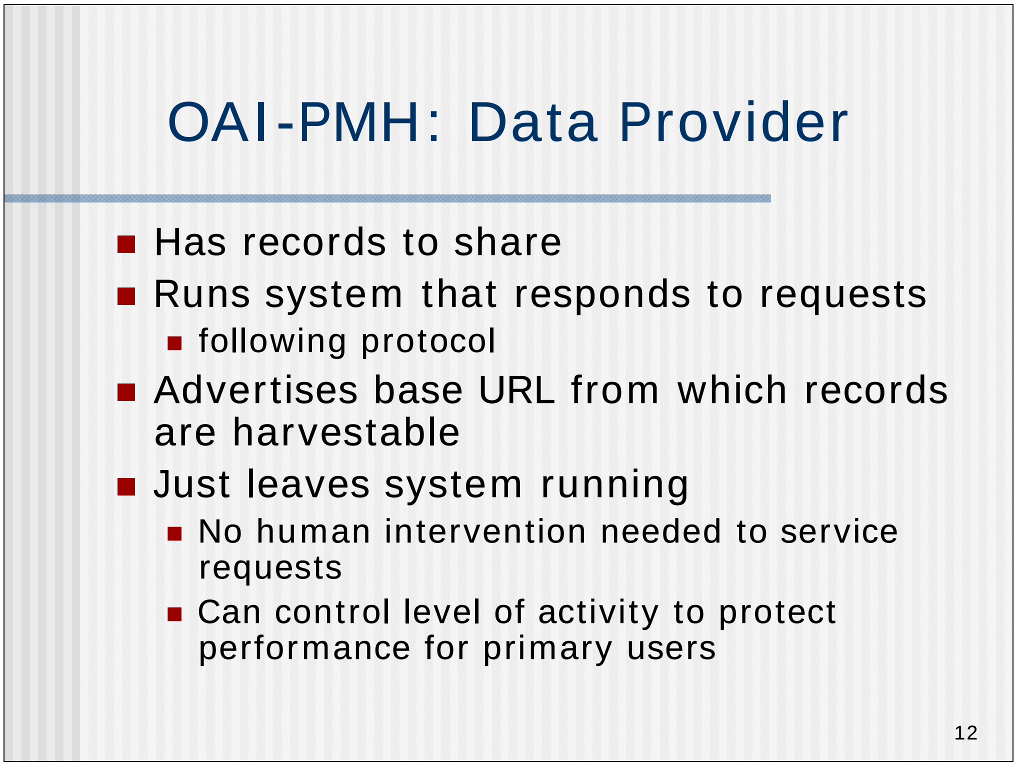
In harvesting metadata, you set the package you’re looking at. In crawling, you’re going out on the open Web with the intent to collect certain things. Harvesting has an inherent permission. The harvester just has to have a harvesting program.

There’s a registry of OAI data providers with a listing of what is there. Sometimes there is a description, sometimes there is not much.

An “item” is the lowest granularity of a resource. In a Word document with two different drafts, the item level could include both drafts, or each one would be a separate item. The data provider decides.

Flow control: if you have thousands of records and don’t want to have to harvest them in one chunk, define the chunks and use them in chunk-order. Get them in small enough chunks that it is easy to manage.

You can’t control how people use your data after harvesting. You can specify what they are allowed to do with them (for example, a rights statement about no commercial use), but there is no enforcement of that. Sometimes a data provider does not even keep track of who uses the metadata.



The system responds to requests:

There is the Data provider and the Service provider

A data provider has metadata to offer.

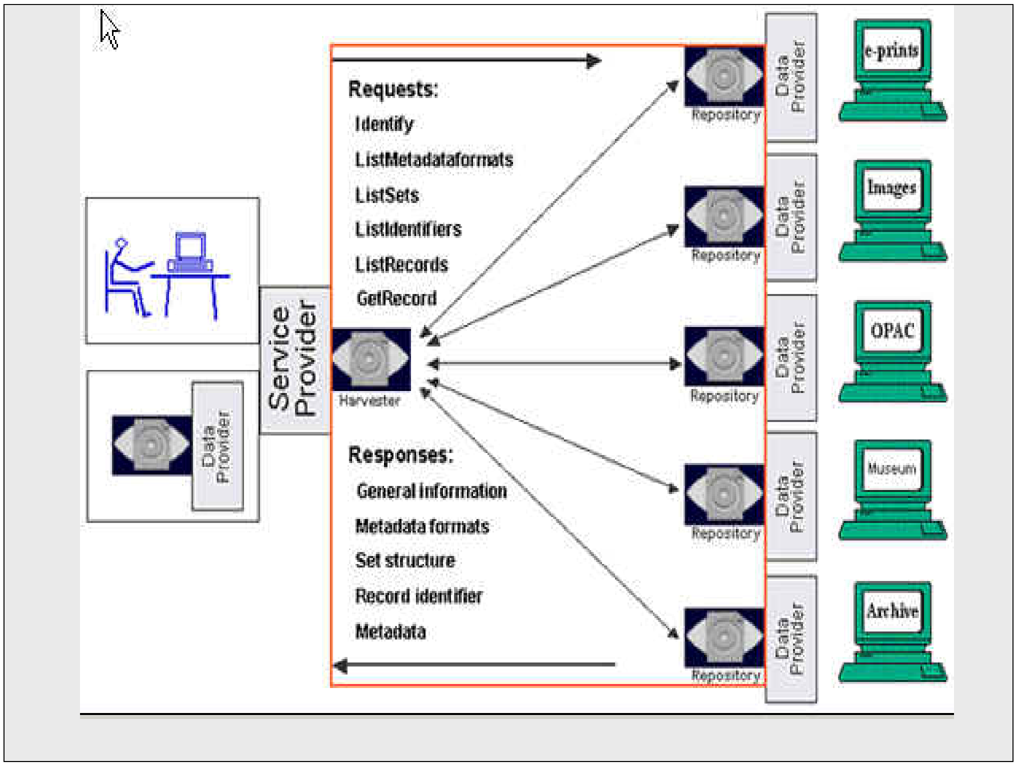
OAI-PMH: Service Provider

- 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)

13

DATA PROVIDER ←=======→ SERVICE PROVIDER

Effort occurs at two poles—OAI seeks to mediate in the middle, with reasonable expectations at both ends.



OAI-PMH: overview and structure model as provided in a tutorial by Leona Carpenter of the Univ. of Bath.

This shows the sorts of requests handled by OAI-PMH and how it works across metadata repositories.

OAI Best Practices Activities

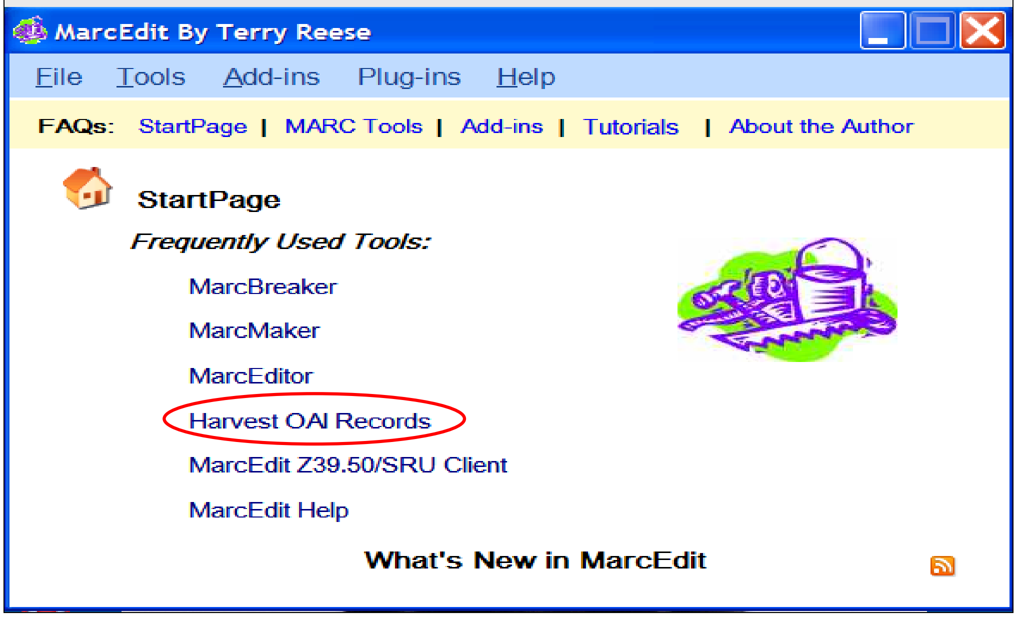
- Sponsored by Digital Library Federation (DLF)
- Guidelines for data providers and service providers
  - http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl
  - Not just DLF, also NSDL
  - Best Practices for Shareable Metadata
    - http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl?PublicTOC
- Workshops to encourage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

15

NSDL=National Science Digital Library

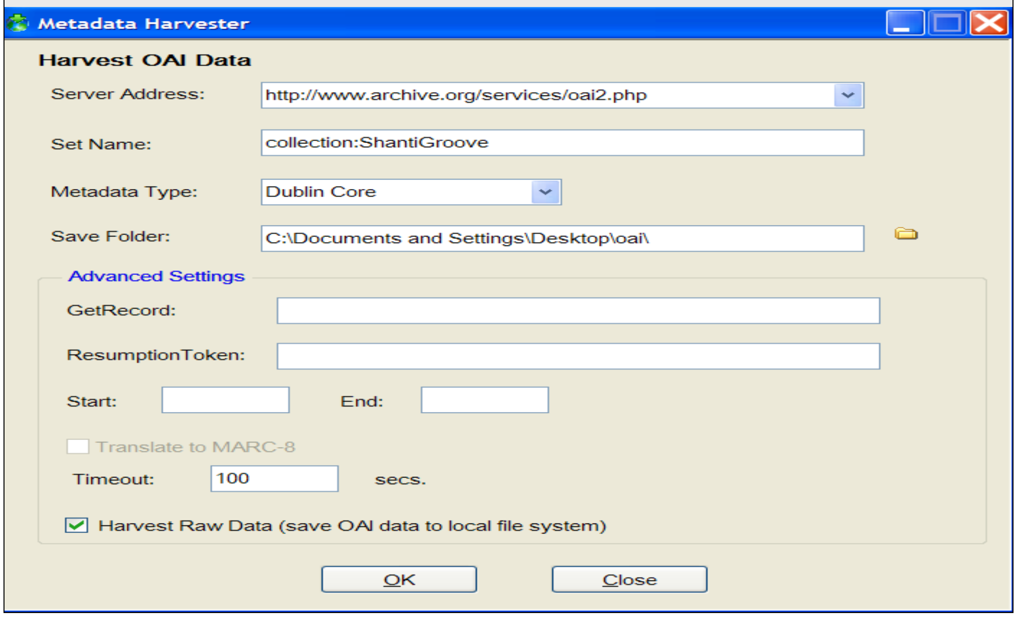
IMLS=Institute of Museum and Library Services

OAI Example



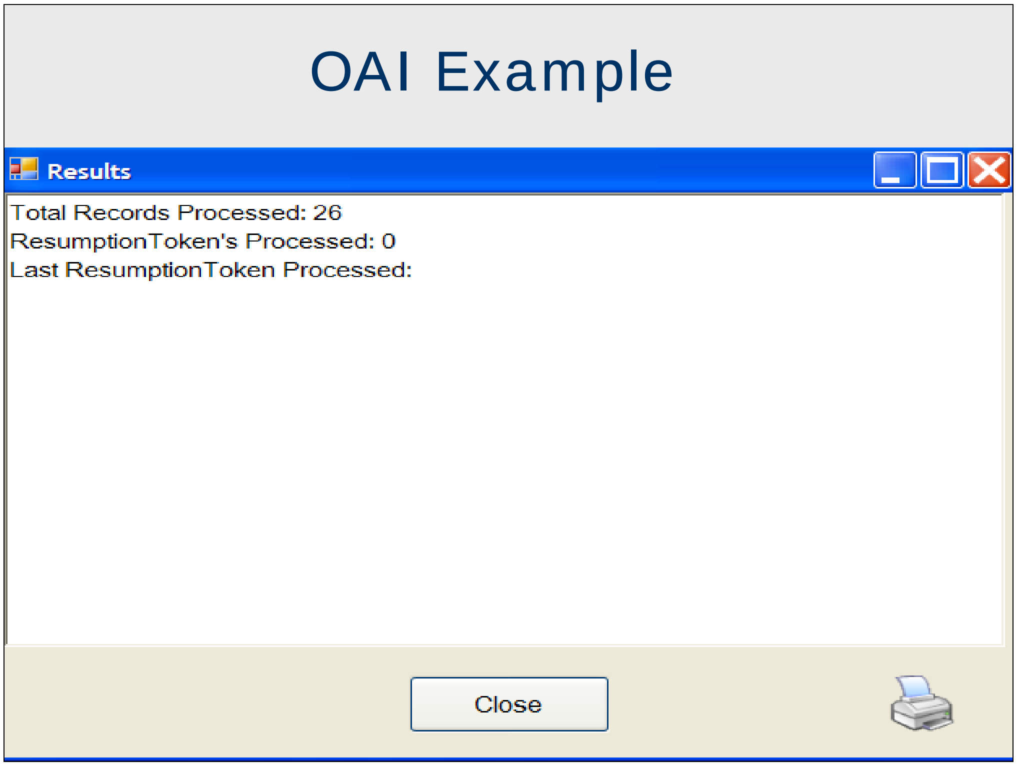
You can use the MarcEdit program to do OAI harvesting. There are numerous other OAI harvesters out there that you can find.

OAI Example



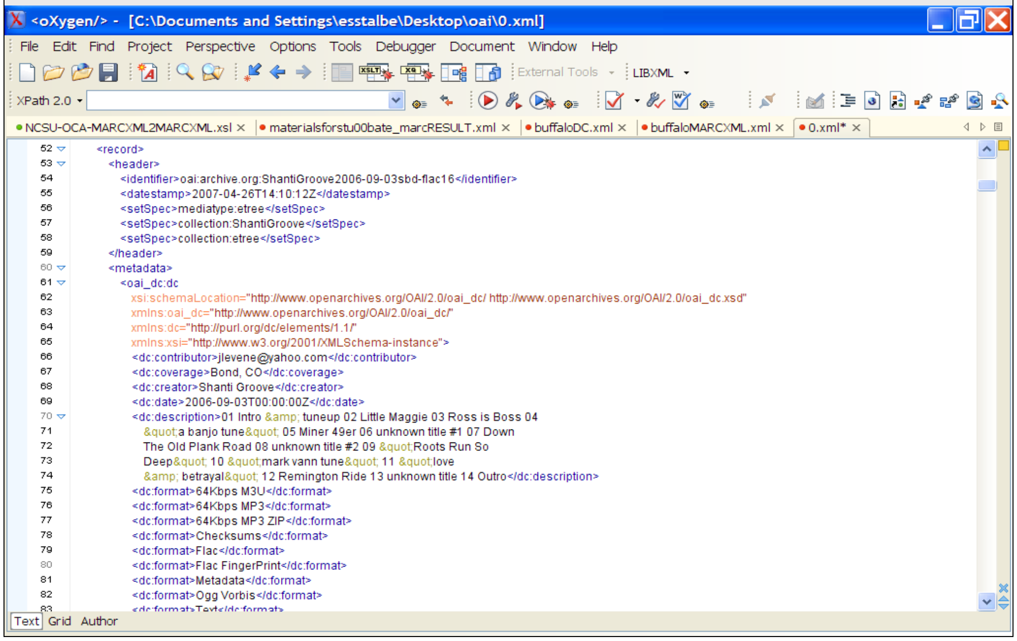
Choose Advanced Setting

1) Check box Harvest Raw Data (do this first or the "save folder" box does not appear)
2) Enter server address
3) Metadata type: choose DC, but show others
4) Save folder on desktop
5) Explain GetRecord – if you wanted a particular record, identify the record number
6) Explain Resumption token, helps you if your download fails midstream to know where it stopped. Leave blank b/c the record set is small
7) Explain Start/End – If you were downloading only a certain portion of the records, this is the start/end record numbers you want to download.

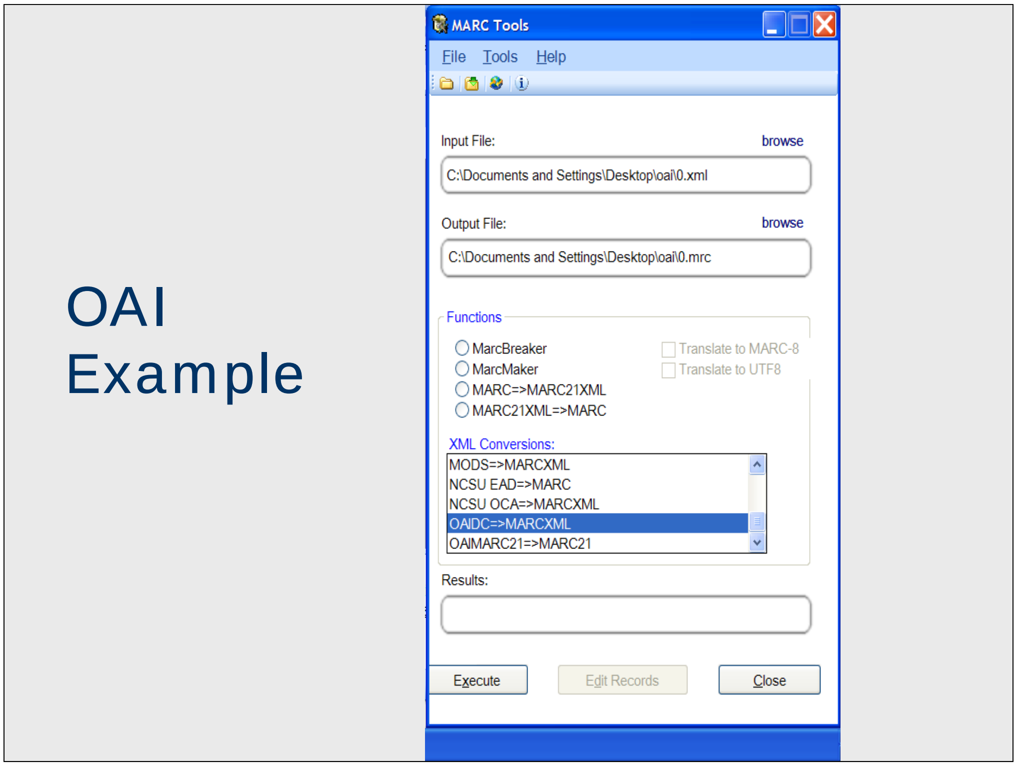


Success! It downloaded 26 records

OAI Example



Open file & look at DC metadata.



Point out that MarcEdit has an OAIDC-MARC stylesheet if you want to turn that file into MARC records. If there is time, you can do that live & show.

OAIster

- 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.
- http://www.oaister.org/

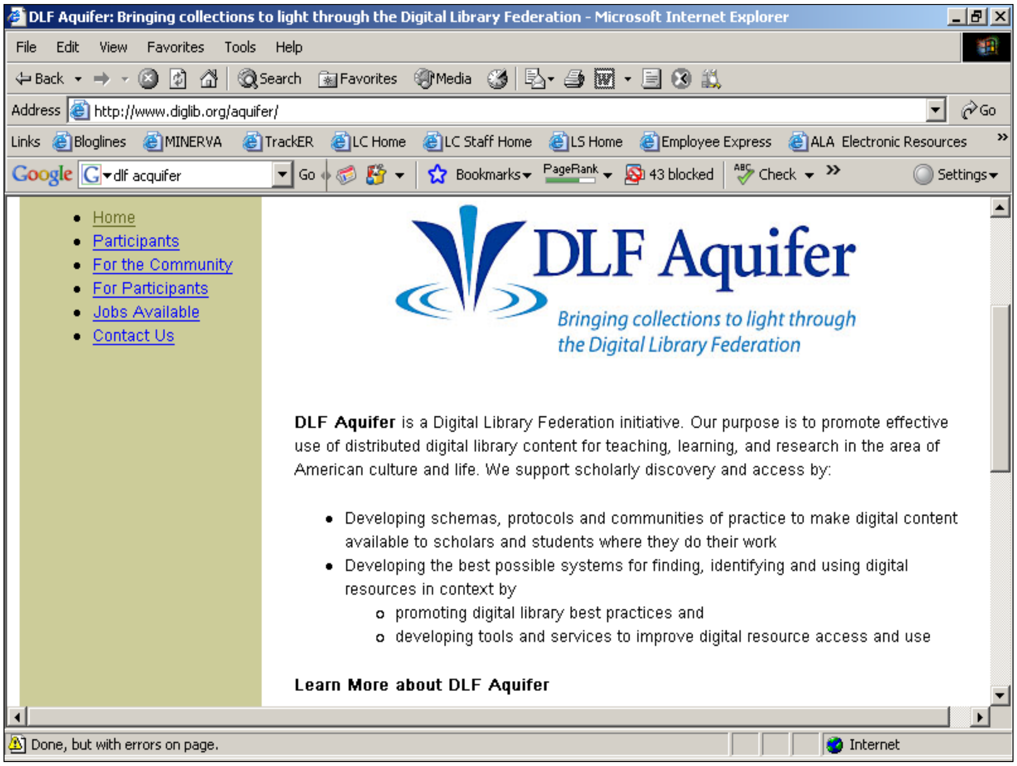
21

This is the homepage for OAIster.

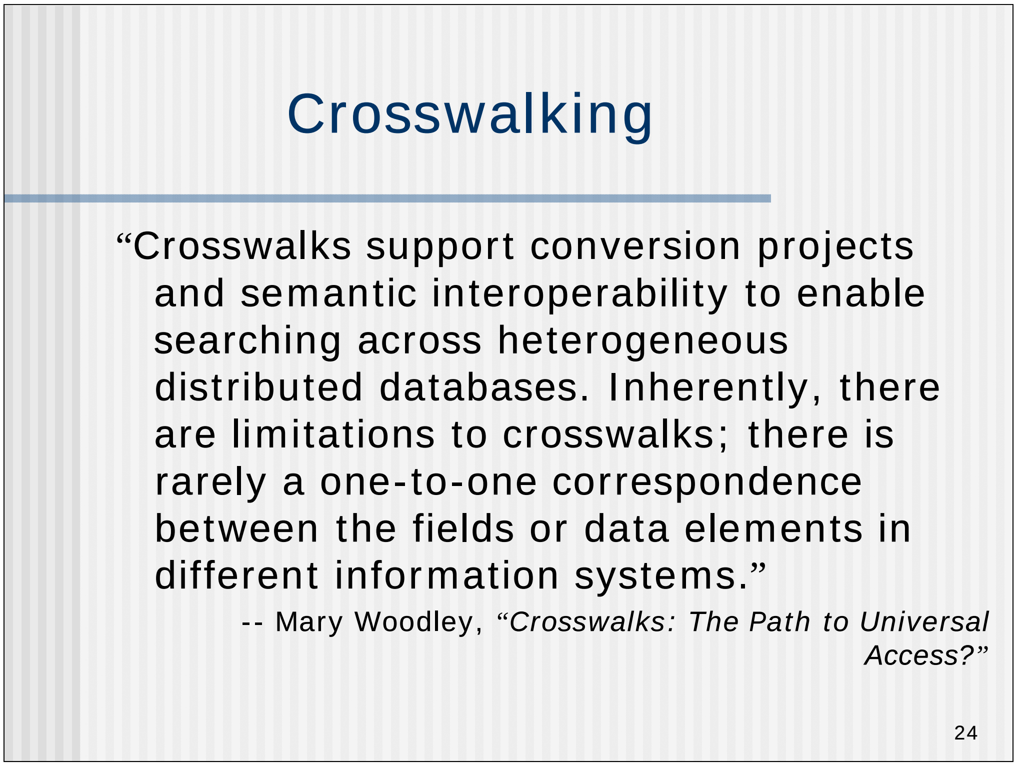
OAIster currently provides access to 15,494,385 records from 939 contributors.

It can be searched by Title, Author/Creator, Subject, Language or Entire Record. Searches can also be limited by resource type (text, image, audio, video, dataset) and sorted by title, author, date and hit frequency. Results allow further limiting by data contributor (i.e., where the record was harvested from).

Go live and do a search. Click on search, then fill in hair as subject and limit to image. Look at the results and see how records differ in the values that appear in the elements. For instance, there are different kinds of values in the element ResourceType: in one record there is “73.57.10.jpg “, in another “1 lock of hair”, in another “
 “ (others have rich metadata that come from controlled vocabularies). This is a result of mapping disparate data to a common field.



The project, *DLF Aquifer Development for Interoperability Across Scholarly Repositories: American Social History Online*, is intended to enable scholars to use distributed digital collections as one in a variety of local environments. The project addresses the difficulty humanities and social science scholars face in finding and using digital materials located in a variety of environments with a bewildering array of interfaces, access protocols and usage requirements. Its focus is digital library collections pertaining to nineteenth and twentieth century United States social history across institutional boundaries.



The vast majority of metadata comes down to beginning with the form you have it in and what form you want to turn it into.

Mapping (conceptual) ←===→ transformation (practice, albeit in an automated way)

At least map from the local (usually richer) scheme to simple DC.

There are various attempts to have lists of crosswalks. Many are not being maintained and versioned.

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

25

Crosswalks require constant maintenance considering that metadata standards are continuously revised.

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

26

Results may be unexpected when converting from one format to the next because of the differences between metadata formats. These differences are…

Example:Mapping MODS:title to DC:title

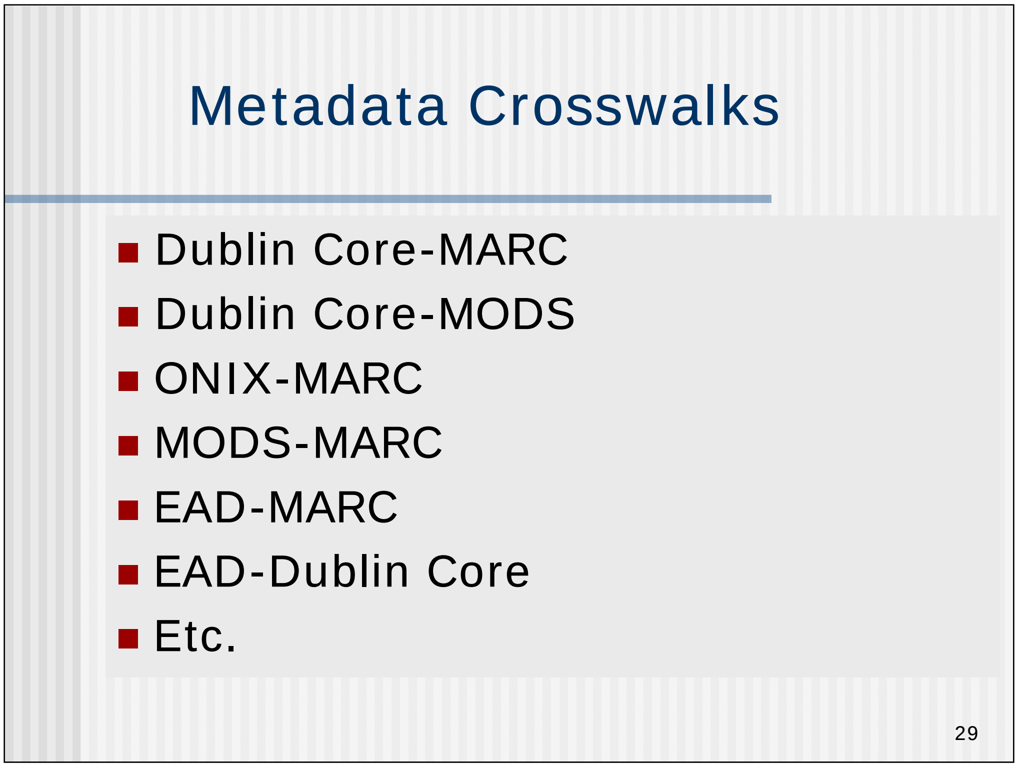
- Includes attribute for type of title
  - Abbreviated
  - Translated
  - Alternative
  - Uniform
- Other attributes: ID, authority, displayLabel, xLink
- Subelements: title, partName, partNumber, nonSort
- Title definition reused by: Subject, Related Item

27

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

28



There are conversion tools, which are usually XSLT stylesheets. LC provides some for converting between MARC and MODS to and from various other formats

There are beginning efforts to maintain crosswalks in registries, which could provide machine readability and facilitate in keeping them updated.

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

30

MARC to Dublin Core Crosswalk (Unqualified)

- Conventions:
    - 1."$" is used to specify the subfield used. If none is specified, use all subfields.
    - 2. DC element is repeated if multiple MARC fields are used.

MARC fields	DC Element	Implementation Notes
100, 110, 111, 700, 710, 711	Contributor	
720		
651, 662	Coverage	
751, 752		
	Creator	Creator element not used.
008/07-10	Date	
260cg		
500-599, except 506, 530, 540, 546	Description	
340	Format	
856$q		
020$a, 022$a, 024$a	Identifier	
856$u		
008/35-37	Language	
041abdefghj		
546		
260ab	Publisher	
530, 760-787ot	Relation	
506, 540	Rights	
534$t	Source	
786ot		
050, 060, 080, 082	Subject	
600, 610, 611, 630, 650, 653		
245, 246	Title	Repeat dc:title for each. Some applications may wish to include 210, 222, 240, 242, 243, and 247.
Leader06, Leader07	Type	See Appendix 2 for Leader-Type rules.
655		

This crosswalk and the one following are from the MARC Website.

This slide shows a crosswalk for unqualified Dublin Core; the following slide is for Qualified Dublin Core. What do you lose, going from MARC to DC? (Example, Creator: some 700s are for related records. You could have a creator value on this record for someone who is actually the creator on another resource.)

This crosswalk tells people to use Source. Source is a specific kind of relationship, so you should probably use Relation instead.

Note that the crosswalks don't' tell you how to put values into the other metadata scheme, for example, to put parts of a subject into different subfields.

MARC to DC Qualified

http://www.loc.gov/marc/marc2dc.html#qualifiedlist

III. MARC to Dublin Core Crosswalk (Qualified)

MARC fields	DC Element	DC Qualifier(s)	Implementation Notes
541$c	Accrual Method		
310$a	Accrual Periodicity		
521	Audience		
100, 110, 111, 700, 710, 711$e	Contributor	Value in $e.	Roles may be used as refinements of Contributor if using qualified DC. See "Roles" note below.
720$e			
255, 034	Coverage	Spatial	Some 255 information equivalent to DC encoding scheme but different syntax.
522			
650$z, 651, 662			
751, 752			
043$c,044$c	Coverage	Spatial ISO3166	

This includes a portion of the MARC to DC Qualified crosswalk. Bring it up online to show the rest.

With Qualified Dublin Core, we've gained some things but it still does not address a lot of fields.

MARC relator terms may be used as refinements for Contributor. Creator is not used in this mapping because of the difference between the semantics in the two metadata schemes (i.e. not all names in MARC 1XX and 7XX are considered creators and creators can be found in 7XX so there is not a perfect mapping).

DC Qualified allows for stating the source of the vocabulary if controlled. There is still data from the richer MARC record which does not get converted and some data in DC Qualified which is not in MARC.

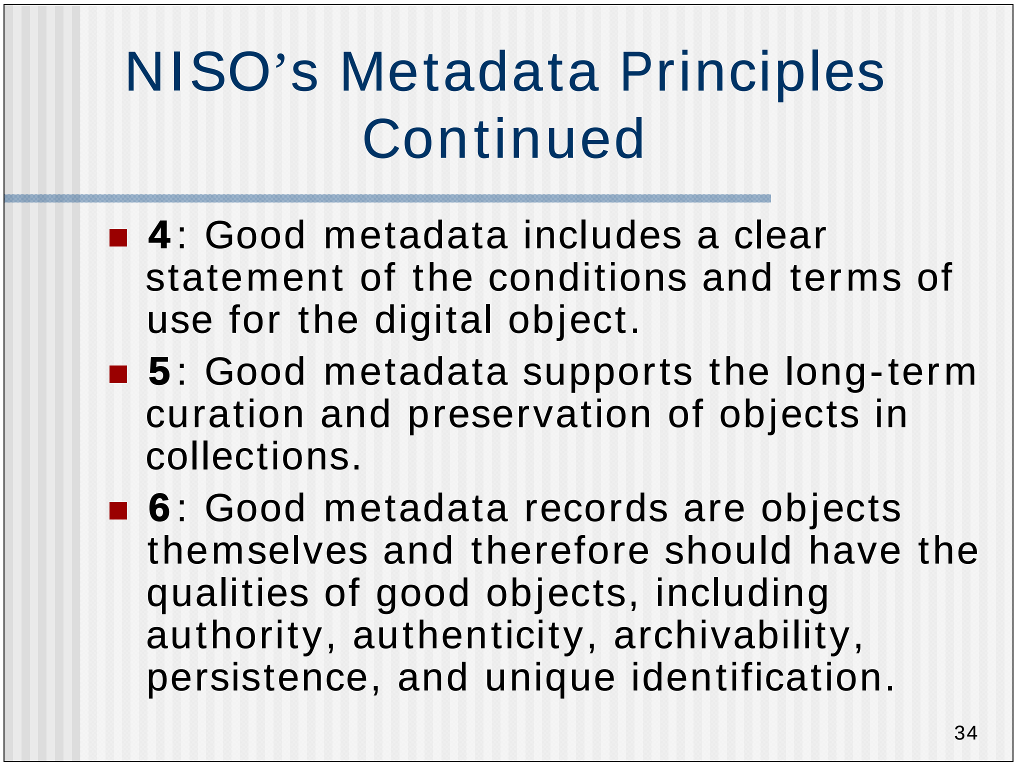
There is a mapping of the MARC Leader/06 (Type of record) values to DC Resource Type, although again there is not a one-to-one correlation and some data will not be accurate.

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

33

Whether mapping, converting, harvesting, or doing original metadata creation, these are good practices for metadata in general.



Libraries have a long history of thinking of metadata use over time. The need to document standards and practices has probably come up already, repeatedly.

Use vocabularies consistently and document them.

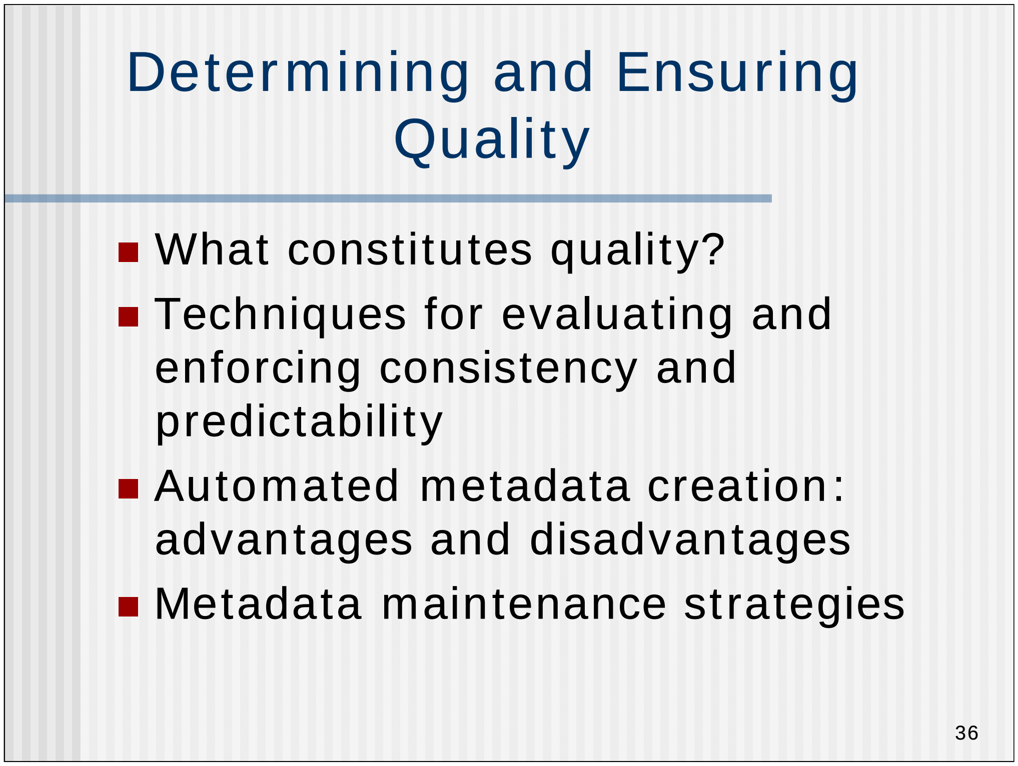
Quality issues

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

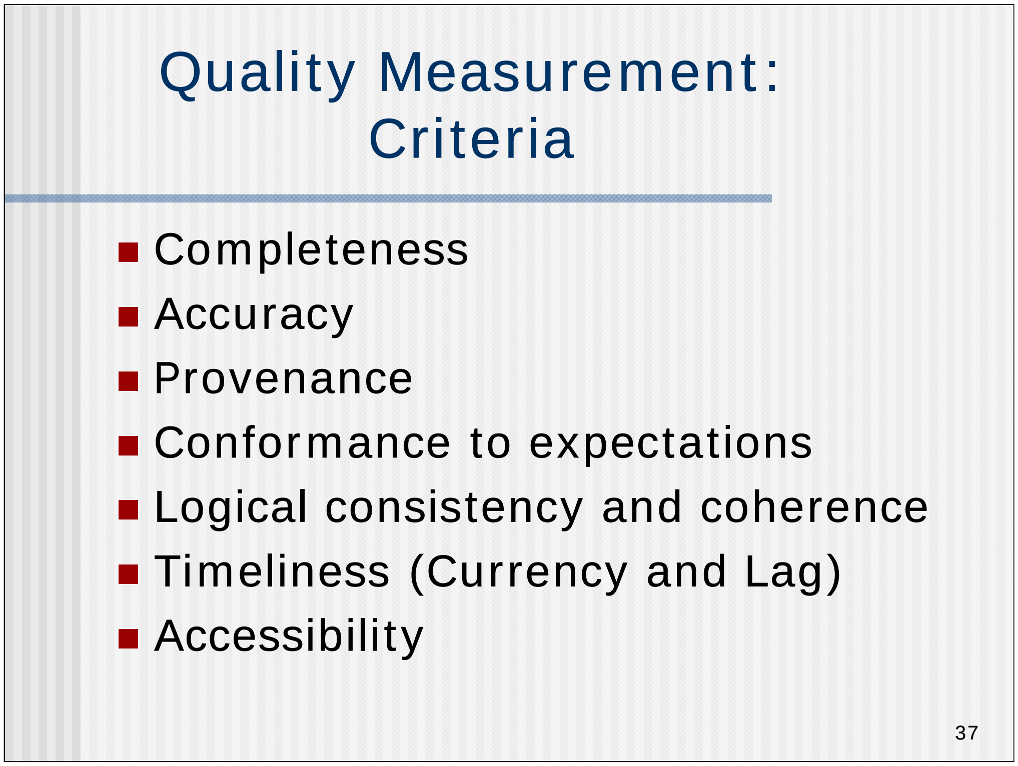
35

Now we will move to issues of quality of metadata.

Quality issues are important in the effective use of metadata. We will consider the following…



Quality in the metadata world may be different than in traditional libraries. It's important that we re-start the conversation about quality with these differences in mind. We want to talk about …



Completeness: describe the item as completely as is economically feasible

Accuracy: apply standards consistently with minimum of error

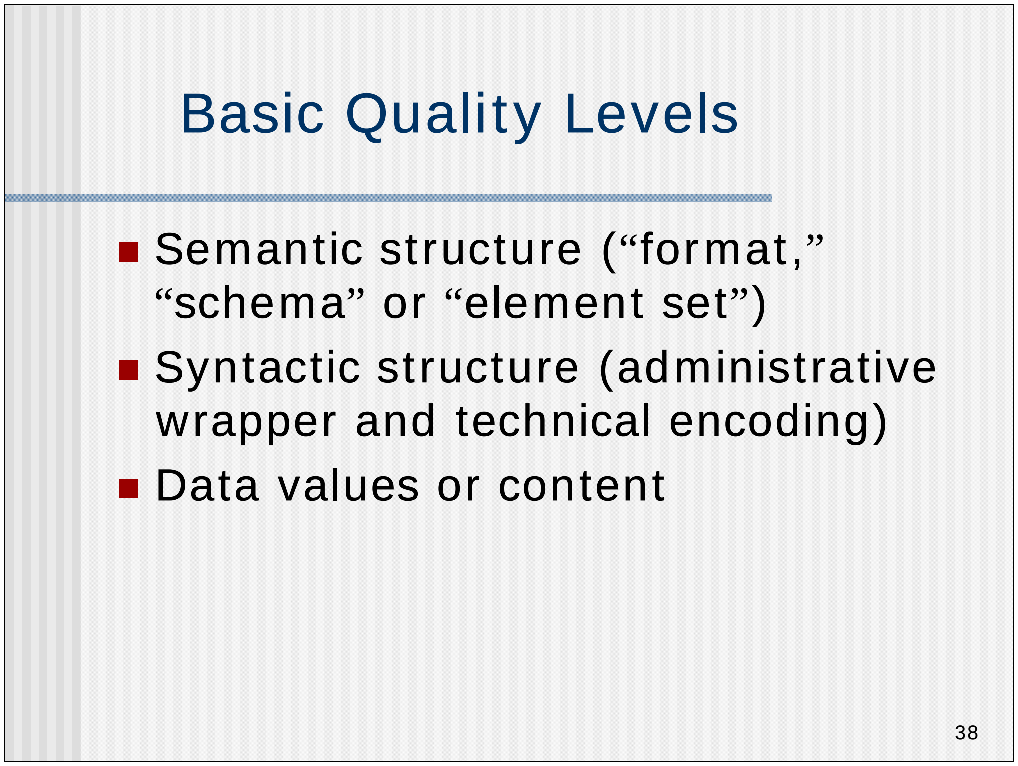
Provenance: provide information on source of metadata, whether it was generated or converted

Conformance to expectations: reflects what the given community would expect, using appropriate elements and controlled vocabulary

Logical consistency: uses application profiles and similar structures

Timeliness: when the target item changes the metadata may need revision

Accessibility: make sure metadata is openly available (if possible) and links are current; should be accessible by people with disabilities



Each of these levels can be evaluated for quality.

One level is semantic structure, meaning that the element set is being appropriately applied in terms of the element definitions.

Another is that it is encoded according to the rules of the syntax (e.g. able to be validated against an XML schema).

Another level is that data values are expressed appropriately, which in the case of controlled vocabularies means that the value could be validated against a machine readable form of that controlled vocabulary.

People are developing tools to analyze quality.

http://www.spotfire.com/

Spotfire Decision Site –a statistical visualization package

You can use such applications to decide to do something with specific collections

“Analyzing Metadata for Effective Use and Re-Use” by Naomi Dushay and Diane Hillman recounts a use of this application. Available from Dublin Core Conference Papers <http://www.siderean.com/dc2003/501_Paper24.pdf>

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)

39

This is the bottom tier of what you should have.

It is desirable to have some control over the metadata itself, including a unique ID for the metadata itself, which is not the same as the Identifier for the object.

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?

40

This is a step up. You know which value is from which vocabulary and are using it properly.

Examples of tokens are: LCSH, AAT, etc.

At this stage, you will assure that your data “plays well with other data” in regular aggregation.

“On the Horse” problem: In the Theodore Roosevelt collection, none of the objects actually had the name of Theodore Roosevelt in the title. Instead, they had titles, like “On the horse” because of course they were all about T.R. The context was lost. This happens all the time!

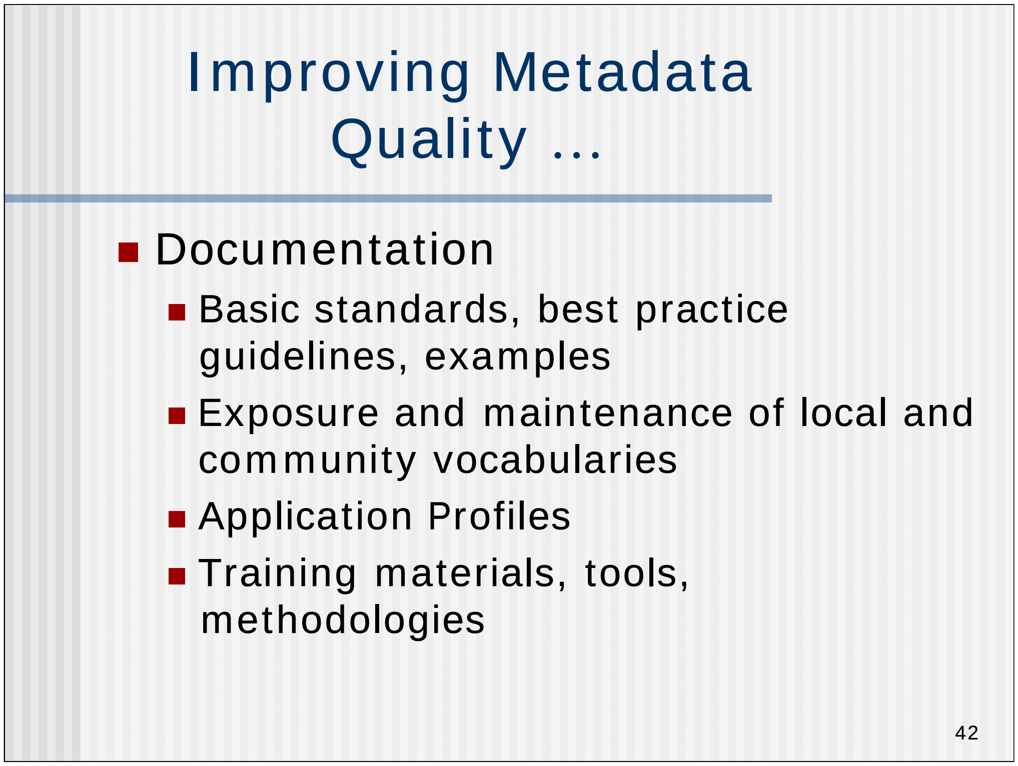
At this stage, there is come indication of the methodology of how the metadata was created: by machine or by a person.

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

41

This level has not been implemented widely, although there is some effort in the METS community to establish METS profiles and for a project to rely on them and perhaps even validate against them. There is also work in the DCMI community to establish application profiles to document practices for high quality metadata.



Question is about how to provide information to people who are using and creating your data.

We are used to making assumptions in the cataloging community about quality, such as the use of content standards like AACR2, etc. This is particularly relevant because of the control we have had over library catalogs.

However in the metadata world– you cannot make those assumptions. As libraries begin to need to process metadata from disparate sources they are having to be more flexible in outlook. Not all those sources will have high quality metadata. As we are initiating digital projects, we need to be concerned about quality of metadata and document what we are using.

Exercise

- Evaluate a small set of machine- and human-created metadata

43

This can be done as an exercise in smaller groups. If desired the alternative exercise may be substituted. See instructor's notes for more detailed information.

Exercise: Evaluate a small set of human and machine-created metadata. The prepared two page handout should be augmented with the printouts of the associated Webpages, which are also included in the packet (these may be eliminated if access to the web is provided in the classroom).

Ages -- machine created (From Identifying Trees)

DC.dot -- machine created (From NSDL metadata registry)

Brazilliance -- human created, with transformations (no resource– all you could see was the record cover)

ICSD Science Zone -- human created

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

SECTION 8: EXERCISES

Evaluate a small set of human and machine-created metadata

More Information

NSDL OAI identifier	oai:infomine-nsdl.ucr.edu:207689
Title	Ages
Subject Keyword(s)	Forest management -- United States, Forests and forestry -- United States, Branching patterns, Distinguishing characteristics, Needles borne, Terminal buds, Tree identification card, Winter key, Trees in winter, Tree identification, Terminal bud, Identifying trees
Description	Objective: Children will learn several features to observe in order to identify trees in winter. Children will identify several trees common to our area based on careful observation of those features.
Link	http://www.nbtc.cornell.edu/mainstreetscience/grab_and_go/treeswinterpg.htm
HTML Title	4-H Grab and Go with Science

DCdot

Dublin Core metadata editor

UKOLN

Results for **URL**: http://purl.org/nsdlregistry/ [summary]

```
<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="NSDL Metadata Registry — Portal" />
<link rel="DC.creator" href="http://eg2.ischool.washington.edu/registry/author/rjlaundr" />
<meta name="DC.subject" content="Contact; Home; WCAG; Document Actions; New SOA Registry from Infravio; News; Advanced
Search; RDF; 1; Usable in any browser; Goals; More news; SKOS; Links to this content; Powered by Plone, the Open
Source Content Management System; National Science Foundation; Log in; ; NSDL Registry Blog; Navigation; Section 508;
Portal; Accessibility; Skip to navigation; ; Events; NSDL Metadata Registry; Sections; Take the Survey; Personal
tools; NSDL Annual Meeting; Ryan Laundry; National Science Digital Library; Join; Documents; The Registry Blog; Skip
to content; Valid CSS; NSDL Registry Public Wiki; 2005-10-03; Valid XHTML; NSDL Registry Grant Awarded; Site Map" />
<meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2005-11-29" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
<meta name="DC.format" content="text/html" />
<meta name="DC.format" content="4858 bytes" />
<meta name="DC.identifier" scheme="DCTERMS.URI" content="http://purl.org/nsdlregistry/" />
```

[how to use the XHTML description]

Album Title:	Brazilliance Vol. 3
Subtitle:	Bud Shank. arrangements by Laurindo Almeida
Notes:	Bud Shank: Flute
Performer:	Almeida, Laurindo
Composer:	Almeida, Laurindo Monk/Hanigan/Williams Shank, Bud Webster, H. D. L. Gershwin, George
Album back/Liner notes:	http://digital-library.csun.edu/IGRA/images/B1D13b.gif
Track 1:	Side a-1 : Harlem Samba : Laurindo Almeida (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 2:	Side a-2 : North of the Border : Laurindo Almeida (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 3:	Side a-3 : Sunset Baion : Laurindo Almeida (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 4:	Side a-4 : ëRound Midnight : Monk/Hanigan/Williams (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 5:	Side a-5 : Toro Dance : Bud Shank (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 6:	Side b-1 : Serenade for alto : Laurindo Almeida (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 7:	Side b-2 : Xana-Lyn : Bud Shank (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 8:	Side b-3 : Blowing Wild : H. D. L. Webster (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 9:	Side b-4 : Gershwin Prelude : George Gershwin (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Track 10:	Side b-5 : Frio y color : Bud Shank (composer) : Laurindo Almeida (performer) : 2 Guitars; or Flute, Guitar
Record Company:	World Pacific Records
Release Date:	1963
Music Genre:	Jazz
Media:	331/3
Collection:	John Tanno Collection
Serial number:	Rec-46
Box number:	1
Disc number:	13
Identifier:	B1D13
Record Label catalog number:	WP-1425
Matrix No. A:	A-826
Matrix No. B:	B-826

NSDL OAI identifier	oai:nsdl.org:nsdl.nsdl:00126
Title	ICSD Science Zone
Subject Keyword(s)	Science--Study and teaching--New York (State)--Ithaca, General science
Description	Welcome to the Ithaca City School District's Science Zone. This supports science education for students in the Ithaca City School District and the global community. Learn science and let it take you places. Enjoy!
Publisher / Resource Provider	Ithaca City School District
Recource type	text/html, image
Link	http://ithacasciencezone.com
Language	en
Grade Level	Grades Pre-K to 12
HTML Title	ICSD Science Zone

Ages:

8 to 12

HANDOUT: IDENTIFYING TREES IN WINTER

Contributor: Susan Jaquette, Cornell Plantations volunteer

Main idea: Trees have distinguishing characteristics that enable identification even in winter.

Objective: Children will learn several features to observe in order to identify trees in winter. Children will identify several trees common to our area based on careful observation of those features.

Materials:

- Small branches from several deciduous trees to illustrate the different branching patterns and terminal buds.
- Small branches from several evergreen trees to illustrate pine needle bundles, needles borne singly, needles borne on a stem or not on a stem, the scale-like foliage of arborvitae or cedar, and differences among cones.
- Book or chart showing characteristic silhouettes of common trees.
- *Know Your Trees* (4-H Bulletin 85) has a winter key that identifies hardwoods based on twigs and buds. There is an online version of the winter key available at: http://cyrus.bh.cornell.edu/tree/winterkey.htm
- Other tree handbooks for reference.

Motivator: To get at the idea of identification through different features, ask children how they might recognize someone they know who was far away, for example across a field. (You might elicit posture, shape, size, and silhouette.) Then ask how do you recognize someone who is nearby, for example, someone in the group. (More specific features will be elicited ? blue eyes, freckles, curly hair, etc.)

Explain that trees, too, have features and can be recognized by careful attention to their features. Scientists have organized many of these features into what's called a taxonomic key for identifying trees.

Questions: Before you begin the activity, ask the children what features they might look at if you were trying to distinguish one tree from another in winter? Elicit as many characteristics as possible.

Activity:

1. Beforehand, the instructor should locate 4 to 6 nearby trees that will be easy to identify based on their particular characteristics. Trees such as red pine, white pine, hemlock, spruce, sycamore, oak, maple, and ash lend themselves to using the distinguishing characteristics presented to the students in Step 3. For example:

    - The two pines have a different number of needles in their clusters.

    - The hemlock and spruce have different cones, and the hemlock?s needles are on short stems while the spruce needles lack stems.

    - The sycamore lacks a terminal bud and has a zigzag branching pattern while the oak has a terminal bud and an alternate branching pattern.

    - The maple and the ash are the only common local large trees with opposite buds and branches.

2. Post a number on each of the trees selected. Make a tree identification card for each tree. Each card should have the tree's name, the descriptive characteristics needed to identify that tree (as stated, for example, in Step 1), possibly a silhouette sketch of the tree, and any other prominent characteristics. (For example, the ash has stout twigs or the sycamore has peeling bark high on its trunk.)

3. Have the children examine the silhouette poster, the branches and cones looking for differences among them. Show them or help them discover the different branching patterns. Point out branches with terminal buds and those without. Have them count pine needles in bundles and observe needles borne singly on or without stems. Have them describe how the cones differ. Everyone should have hands-on experience and practice at this time.

4. Divide children into as many groups as trees you have numbered. Give each group a tree identification card and instruct them to examine each numbered tree until they are sure they have found the one on their card.

5. After all groups have found their trees, gather the entire group together and go from tree to tree having each group point out the distinguishing characteristics of their tree.

Learning checks: In addition to the built in check in Step 5, instructor can have the group examine other nearby trees and demonstrate their knowledge of branching patterns, terminal buds, and needle arrangement.

Background: Basic knowledge about tree identification and familiarity with distinguishing characteristics of the 4 to 6 trees chosen for the activity.

Vocabulary:

Terminal bud: The bud found at the end of a twig.
Opposite: Arrangement of leaves or buds so that they connect to branches at points directly across from each other.
Alternate: Arrangement of leaves or buds where they are staggered or not placed directly across from each other on the twig.
Zig-zag branches: Change direction at each bud, creating a zig-zag pattern.
Needle cluster or bundle: A group of needles attached together at one point to a twig.
Taxonomic key: A guide that helps you find the name of an organism, usually by asking sets of paired questions.

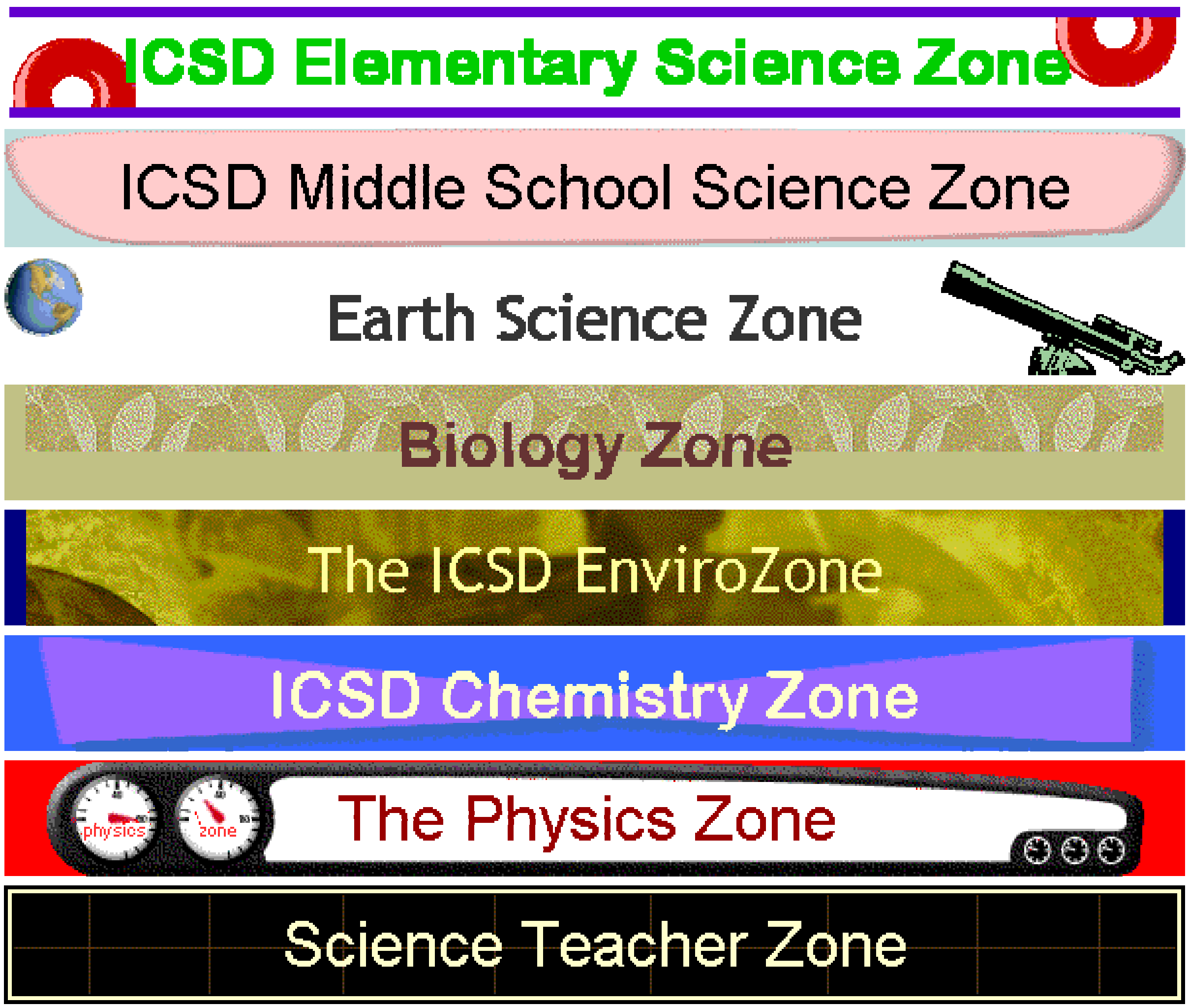
Extensions:

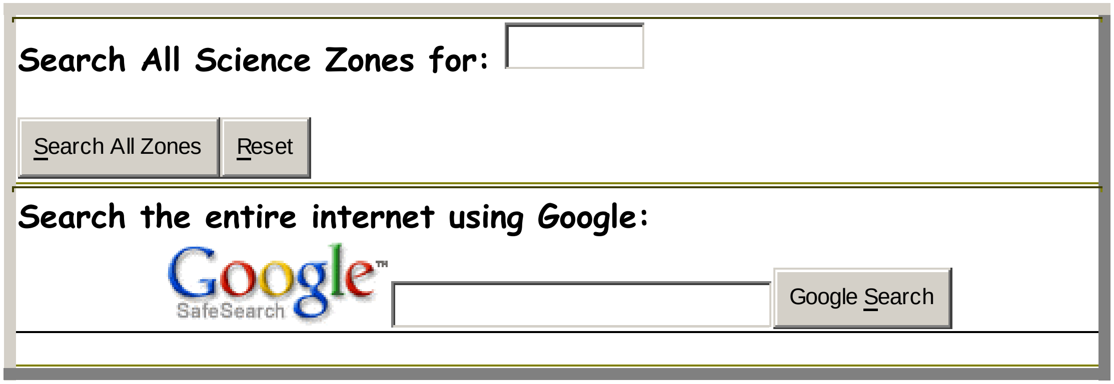
- Have groups switch cards and identify other trees (writing down their answers) before moving on to Step 5.

Number additional "distracter" trees so students must consider and observe more trees.

ICSD SCIENCE ZONE

Welcome to the Ithaca City School District's Science Zone. This supports science education for students in the Ithaca City School District and the global community. Learn science and let it take you places. Enjoy!




Search All Science Zones for:
Search All Zones
Reset
Search the entire internet using Google:
Google
SafeSearch
Google Search

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Dublin Core to MARC Crosswalk

Network Development and MARC Standards Office
Library of Congress

Date issued: 2008-04-23

Previous version: http://www.loc.gov/marc/dccross_20010312.html

I. Introduction

The following is a crosswalk between the metadata terms in the Dublin Core Element Set and MARC 21 bibliographic data elements. The crosswalk may be used for conversion of Dublin Core metadata into MARC, for instance as a tool for developing XSLT transformations. For conversion of MARC 21 into Dublin Core, the MARC to Dublin Core Crosswalk should be used, since in this conversion many fields are mapped into a single Dublin Core element.

In the Dublin Core to MARC mapping, two mappings are provided, one for unqualified Dublin Core elements (i.e. the main fifteen elements as in Dublin Core Metadata Element Set, version 1.1) and the other for qualified (those in addition that appear only in DCMI Metadata Terms). The latter includes both refinements of the original fifteen as well as syntax and vocabulary encoding schemes.

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.

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 label.

II. Dublin Core to MARC Crosswalk (15 elements in version 1.1 and refinements)

Contributor -- An entity responsible for making contributions to the resource.

Unqualified:
720 ##$a (Added Entry--Uncontrolled Name/Name)

Qualified:
Element refinements may be used from the MARC relator list for those specified roles that refine dc:contributor. If DC metadata includes a role refinement (only used for Contributor), use the term in 720$e or the code in 720$t. See: MARC Relator Terms and Dublin Core

Note: there is no way to specify whether the Contributor is a person or organization because it is not in the Dublin Core data. If it can reasonably be determined that the contributor is a person or organization, fields 700 1#$a (Added Entry--Personal Name) or 710 2#$a (Added Entry--Corporate Name) may be used.

Coverage -- The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which he resource is relevant.

Unqualified:

500$a (General note)

Qualified:

Spatial: 522 ##$a (Geographic Coverage Note)

Temporal: 513 ##$b (Type of Report and Period Covered Note/Period covered)

Syntax encoding schemes:

Box: 507$a (Scale Note for Graphic Material)

ISO3166: 043$c with $2 iso3166 (Geographic Area Code/ISO code)

Point: 507$a (Scale Note for Graphic Material)

Vocabulary encoding scheme under spatial:

TGN: 651 #7 $a (Subject Added Entry--Geographic Name) with $2=tgn

Period: 045 0# $b (Time Period of Content/Formatted 9999 B.C. through C.E. time period)

Creator -- An entity primarily responsible for making the resource.

Unqualified:

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

Note: there is no way to specify whether the Creator is a person or organization because it is not in the Dublin Core data. If it can reasonably be determined that the Creator is a person or organization, fields 100 1#$a (Main Entry--Personal Name) or 110 2#$a (Main Entry--Corporate Name) may be used for the first Creator, and 700 1#$a (Added Entry--Personal Name) or 7102#$a (Added Entry--Corporate Name) for others may be used.

Date -- A point or period of time associated with an event in the lifecycle of the resource.

Unqualified:

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

Qualified:

Available: 307 ##$a (Hours, Etc.)

Created: 046$k (Special Coded Dates/Date created)

Date Accepted: 502##$a (Dissertation Note) with initial label "Date accepted"

Date Copyrighted: 260##$c (Date of publication, distribution, etc.) or 542 $g (Information Related to Copyright Status/Copyright date) *Note: this field was defined in MARC in January 2008. Prefer 542 if field is available.*

Date Submitted: 502##$a (Dissertation Note) with initial label "Date submitted"

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

Modified: 046##$j (Special Coded Dates/Date modified)

Valid: 046##$m (Special Coded Dates/Date valid)

Syntax encoding schemes:

Period: 045##c (Special Coded Dates/Date 1, C.E. date)

W3CDTF: 260 ##$c (Date of publication, distribution, etc.); date may also be generated in 008/07-10; see below under Notes. Remove hyphens in 008.

Description -- An account of the resource.

Unqualified:

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

Qualified:

Abstract: 520 ##$a (Summary, etc. note)

TableofContents: 505 0#$a (Formatted Contents Note)

Format -- The file format, physical medium, or dimensions of the resource.

Unqualified:

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

Qualified:

Extent: 300 ##$a (Physical Description)

Medium: 340 ##$a (Physical Medium)

Syntax encoding scheme:

IMT: 856 ##$q (Electronic Location and Access/Electronic Format Type)

Identifier -- An unambiguous reference to the resource within a given context.

Unqualified:

If string begins http://: 856 40 $u (Electronic Location and Access/URI)

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

Qualified:

Bibliographic Citation: 500 ##a (General note) with initial label "Bibliographic citation"

Syntax encoding scheme:

URI: 856 40$u (Electronic Location and Access/Uniform Resource Locator)

Note: other types of identifiers (e.g. ISSN, ISBN) may be expressed in the form of URIs in dc:identifier.

Language -- A language of the resource.

Unqualified:

546 ##$a (Language note)

Qualified:

Syntax encoding schemes:

ISO 639-2: 0410#$a (Language code)

RFC 1766: 041 07$a (Language code) with $2=rfc1766

RFC 3066: 041 07$a (Language code) with $2=rfc3066

RFC 4646: 041 07$a (Language code) with $2=rfc4646

Publisher -- An entity responsible for making the resource available.

Unqualified:

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

Relation -- A related resource.

Unqualified:

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

Qualified:

URI (Encoding scheme): 787 0#$o (Nonspecific Relationship Entry/Other identifier)

Conforms To: 787 0#$n (Nonspecific Relationship Entry/Other identifier) with $i=Conforms to

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)

Is Based On: 786 0#$n (Data Source Entry/Note)

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

Is Referenced By: 510 0#$a (Citation/References Note/Name of source)

Requires: 538 ##$a (System Details Note)

Is Required By: 787 0#$n (Nonspecific Relationship Entry/Other identifier) with $i=Is required by

Replaces: 780 00$n (Preceding entry)

Replaces (Scheme=URI): 780 00$o (Preceding entry)

Is Replaced By: 785 00$n (Succeeding entry)

Is Replaced By (Scheme=URI): 785 00$o (Succeeding entry)

References: 787 0#$n (Nonspecific Relationship Entry/Other identifier) with $i=References

References (Scheme=URI): 787 0#$o (Nonspecific Relationship Entry/Other identifier) with $i=References

Rights -- Information about rights held in and over the resource.

Unqualified:

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

Qualified:

Access Rights: 506##$a (Restrictions on Access Note)

License: 540##$a (Terms Governing Use and Reprofuction Note

License (if value is a URI): 540##$u

Source -- A related resource from which the described resource is derived.

Unqualified:

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

Qualified:

Syntax encoding scheme

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

Subject -- The topic of the resource.

Unqualified:

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

Qualified:

Vocabulary encoding schemes:

DDC: 082 ##$a (Dewey Decimal Call Number/Classification number)

LCC: 050 ##$a (Library of Congress Call Number/Classification number)

LCSH: 650 #0$a (Subject added entry--Topical term)

MeSH: 650 #2$a (Subject added entry--Topical term)

UDC: 080 ##$a (Universal Decimal Classification Number)

Title -- A name given to the resource.

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)

Type -- The nature or genre of the resource.

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

Qualified:
DCMI Type vocabulary (Encoding scheme): 655 #7$a (Index Term--Genre/Form) with $2=dct

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

III. Additional Dublin Core Elements

Accrual Method: 541##$c (Source of Acquisition Note)

Accrual Periodicity: 310##$a (Current Publication Frequency)

Audience: 521##$a (Target Audience Note)

Education Level: 521##$a (Target Audience Note)

Provenance: 561##$a (Ownership and Custodial History)

IV. 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 MARC 21 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 and convert the data elements to MARC 21 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. 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. A primary use for this mapping is for harvesting metadata records from different sources, where there may be a need to translate Dublin Core records to MARC or vice versa.

Go to:

MARC Home Page

Library of Congress Home Page

Library of Congress
Library of Congress Help Desk (04/25/2008)

MARC to Dublin Core Crosswalk

Development and MARC Standards Office
Library of Congress

Date issued: 2008/04/24

Previous version: http://www.loc.gov/marc/marc2dc-2001.html

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 to 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 by 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 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 DCMI Metadata Terms.

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 mapping to more than one DC element, a decision may need to be made which Dublin Core element to map it to.

II. MARC to Dublin Core Crosswalk (Unqualified)

Conventions:

1."$" is used to specify the subfield used. If none is specified, use all subfields.

2. DC element is repeated if multiple MARC fields are used.

MARC fields	**DC Element**	**Implementation Notes**
100, 110, 111, 700, 710, 711	Contributor	
720		
651, 662	Coverage	
751, 752		
	Creator	Creator element not used.
008/07-10	Date	
260cg		
500-599, except 506, 530, 540, 546	Description	
340	Format	
856$q		
020$a, 022$a, 024$a	Identifier	
856$u		
008/35-37	Language	
041abdefghj		
546		
260ab	Publisher	
530, 760-787ot	Relation	
506, 540	Rights	
534$t	Source	
786ot		
050, 060, 080, 082	Subject	
600, 610, 611, 630, 650, 653		
245, 246	Title	Repeat dc:title for each. Some applications may wish to include 210, 222, 240, 242, 243, and 247.
Leader06, Leader07	Type	See Appendix 2 for Leader-Type rules.
655		

III. MARC to Dublin Core Crosswalk (Qualified)

MARC fields	DC Element	DC Qualifier(s)	Implementation Notes
541$c	Accrual Method		
310$a	Accrual Periodicity		
521	Audience		
100, 110, 111, 700, 710, 711$e	Contributor	Value in $e.	Roles may be used as refinements of Contributor if using qualified DC. See "Roles" note below.
720$e			
255, 034	Coverage	Spatial	Some 255 information equivalent to DC encoding scheme but different syntax.
522			
650$z, 651, 662			
751, 752			
043$c,044$c	Coverage	Spatial ISO3166	
651 if ind2=7 & $2=tgn		Spatial TGN	
033$a	Coverage	Temporal	
533$b			
260$c if precedes date	Date	Copyrighted	
542$g			This field was defined in MARC in January 2008.
260cg	Date	Created	
533$d			
008/07-10	Date	Issued	
260$c			
046$j	Date	Modified	
046mn	Date	Valid	
520 if ind1=# or 3	Description	Abstract	
505	Description	TableofContents	

300$a	Format	Extent	
533$e			
856$q		IMT	
340$a		Medium	
856$u	Identifier	URI	
008/35-37	Language	IS0369-2	
041 with no $2		ISO639-2	
041 with $2=iso639-3		ISO639-3	
041 with $2=rfc1766		RFC1766	
041 with $2=rfc3066		RFC3066	
041 with $2=rfc4646		RFC4646	
561	Provenance		
530	Relation	HasFormat	
776nt			
530$u	Relation	HasFormat URI	
776$o			
774nt	Relation	HasPart	
774$o	Relation	HasPart URI	
775nt	Relation	HasVersion	
775$o	Relation	HasVersion URI	
530	Relation	IsFormatOf	
776nt			
530$u	Relation	IsFormatOf URI	
776$o			
440, 490, 800, 810, 811, 830	Relation	IsPartOf	
760, 773nt			
760, 773$o	Relation	IsPartOf	

		URI	
510	Relation	IsReferencedBy	
785nt	Relation	IsReplacedBy	
785$o	Relation	IsReplacedBy URI	
775,786nt	Relation	IsVersionOf	
775, 786$o	Relation	IsVersionOf URI	
780nt	Relation	Replaces	
780$o	Relation	Replaces URI	
538	Relation	Requires	
506ad	Rights	AccessRights	
540ad			
542$d	Rights	RightsHolder	This field was defined in MARC in January 2008.
786$o	Source	URI	
082	Subject	DDC	
050	Subject	LCC	
600, 610, 611, 630, 650 if ind2=0	Subject	LCSH	
600, 610, 611, 630, 650 if ind2=2	Subject	MeSH	
060	Subject	NLM	
655 if ind2=7 & $2=tgn	Subject	TGN	
080	Subject	UDC	
245	Title		
130, 210, 240, 242, 246, 730, 740	Title	Alternative	
Leader06, Leader07	Type	DCMI Type Vocabulary	See Appendix 2 for Leader-Type rules.
655 if $2=dct			

Appendix 1--Notes

DC Refinements and encoding schemes. Some DC refinements or encoding schemes have not been included because they are not generally found or can not be specifically identified in MARC data. Examples include DCMI Point, DCMI Box and Mediator.

Roles. Those roles in the MARC Relators list that indicate that they refine "Contributor" may be used in qualified DC. See: Relator Terms and Dublin Core Elements.

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

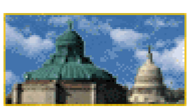
Multiple Type fields may be used; conversions below may result in 2 or 3 Type fields. There are several additional sources of type information in the MARC record; only coded values in Leader/06 and Leaader/07 are detailed in this chart. Field 655 may also be used for more specific type information.

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 (05/02/2008)

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

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

Baker, Thomas 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

Bibliographic Control of Web Resources: A Library of Congress Action Plan http://www.loc.gov/catdir/bibcontrol/actionplan.pdf

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

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

DCMI Initiative. Available at: http://dublincore.org/

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

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

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

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

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

Godby, Carol Jean, 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

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

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

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

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

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

Metadata Encoding & Transmission Standard (METS). Available at: http://www.loc.gov/standards/mets/
Metadata Encoding and Transmission Standard: Primer and Reference Manual http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf

Metadata for Images in XML Schema (MIX): http://www.loc.gov/mix

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

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

Metadata Object Description Schema (MODS). Available at: http://www.loc.gov/standards/mods/
MODS User Guidelines, vers. 3 http://www.loc.gov/standards/mods/v3/mods-userguide.html

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

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

National Information Standards Organization. Understanding Metadata. http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

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

Open Archives Initiative http://openarchives.org

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

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

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

Resource Description and Access (RDA):
http://www.collectionscanada.gc.ca/jsc/index.html

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

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

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/

St. Pierre, Margaret 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

Texas Digital Library Metadata Working Group. MODS Application Profile for Electronic Theses and Dissertations. Version 1. December 2005
http://metalogger.files.wordpress.com/2007/06/tdl-etd-mods-profile.pdf

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

University of Maryland Descriptive Metadata Tag Library
http://www.lib.umd.edu/dcr/publications/taglibrary/umdm.html

UVa Metadata
http://lib.virginia.edu/digital/metadata/

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

Using Metadata Standards in Digital Libraries: Implementing 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/

World Digital Library
http://www.worlddigitallibrary.org/project/english/index.html

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

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

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

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/)

THIS PAGE INTENTIONALLY LEFT BLANK
FOR DOUBLE SIDED COPY

Handout: XML Definitions and Genealogy

Definitions:

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.*

XML Schema Genealogy:

SGML	*Standard Generalized Markup Language, invented in the 1970's, became ISO 8879 in 1986. Very powerful, used extensively in the military and government, as well as for markup for narrative structures (see Text Encoding Initiative for more detail).*
HTML	*Hypertext Markup Language, an SGML application. Restricted set of tags, designed to describe web pages for purposes of presentation. Current version is HTML 4.0.*
XML	*Extensible Markup Language, originally intended as a 'lite' version of SGML, first version unveiled in 1998.*
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.*
XSLT	*Extensible Stylesheet Language Transformations, a general purpose language for transforming one XML document into another, sometimes (not always) for web page display.*

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.*

Exercise for Session 2: Descriptive Metadata
Instructors' information and Answer sheet

- Each group will be given a printout of a digital object
- Create a brief metadata record based on the standard assigned to your group (MODS or DC)
- 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 printout of different digital objects (selected by trainer).
Two examples (provided):

Example 1:
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

Example 2:
52nd Street, New York, N.Y.
http://lcweb2.loc.gov/diglib/ihas/loc.natlib.gottlieb.02771/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
URL

Answers are given in the Metadata Template below:

MODS metadata: example 1 (sheet music)

Data Element (Field)	**Data Value (Content)**	**Controlled Vocabulary (Yes / No)** (please specify)
titleInfo title	America's pinch hit march	
titleInfo subTitle	the hit that ended the worlds greatest war	
name namePart	Dempsey, Bertha Stanfield Type="personal" Authority="lcnaf" (optional)	
name role	composer	Yes: MARC relators
typeOfResource	notated music	Yes (MODS)
originInfo place	Joplin, Missouri	
originInfo publisher	The McMillan Music Co.	
originInfo dateIssued	1919	

Data Element (Field)	**Data Value (Content)**	**Controlled Vocabulary (Yes / No)** (please specify)
PhysicalDescription	3 p.	
subject topic	Baseball—Songs and music	Yes: LCSH
location/url	http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200033287/default.html	
Also useful: RelatedItem type="host" Title	Baseball sheet music collection	

Dublin Core metadata: example 1

Data Element (Field)	**Data Value (Content)**	**Controlled Vocabulary (Yes / No)** (please specify)
Title	America's pinch hit march : the hit that ended the worlds greatest war	
Creator	Bertha Stanfield Dempsey	
Resource Type	Image	Yes: DCMI type
Publisher	The McMillan Music Co.	
Date	1919	
Format	3 p.	
Subject	Baseball—Songs and music	Yes: LCSH
Identifier	http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200033287/default.html	
Also useful: Relation	Baseball sheet music collection	

MODS metadata: example 2 (image)

Data Element (Field)	**Data Value (Content)**	**Controlled Vocabulary (Yes / No) (please specify)**
titleInfo title	52^{nd} Street, New York, N.Y., ca. 1948	
titleInfo subtitle		
Name type="personal" NamePart	Gottlieb, William P.	
name namePart type="date"	1917-	
name role	photographer	Yes: MARC relator
typeOfResource	still image	Yes (MODS)
originInfo place	New York	

Data Element (Field)	Data Value (Content)	Controlled Vocabulary (Yes / No) (please specify)
originInfo publisher		
originInfo dateCreated	1948	
physicalDescription form	graphic	Yes (gmd)
physicalDescription extent	1 negative : color ; 4 x 5 in.	
subject geographic	New York (N.Y.)	Yes (lcsh)
subject topic temporal		
Location/url	http://lcweb2.loc.gov/diglib/ihas/loc.natlib.gottlieb.02771/default.html	
Also useful: accessCondition	Original negative not served.	

Dublin Core metadata: example 2

Data Element (Field)	Data Value (Content)	Controlled Vocabulary (Yes / No) (please specify)
Title	52nd Street, New York, N.Y., ca. 1948	
Creator	Gottlieb, William P.	
Resource Type	image	Yes: DCMI type
Publisher		
Date	1948	
Format	image/jpeg	Yes: (MIME)
Subject	New York (N.Y.)	
Identifier	http://lcweb2.loc.gov/diglib/ihas/loc.natlib.gottlieb.02771/default.html	
Also useful: Rights	Original negative not served.	

Session 3: Administrative metadata (PREMIS)

Example 1: America's pinch hit march

We will only fill in a template for one of the files. A complete example would include a record for each file, showing the relationship with the relationship element. For this example we would have 3 records, one for each file (representing each page).

Data Element (Field)	**Data Value (Content)**	**Controlled Vocabulary (Yes / No)** *(please specify)*
objectIdentifier objectIdentifierType	URI	yes
objectIdentifier objectIdentifierValue	http://lcweb2.loc.gov/natlib/ihas/service/encyclopedia/200033287/0001v.jpg	
objectCharacteristics size	629507	
objectCharacteristics format formatDesignation formatName	image/jpeg	yes
objectCharacteristics format formatDesignation formatVersion		
creatingApplication creatingApplicationName	Adobe photoshop	
creatingApplication creatingApplicationVersion	version CS3	
creatingApplication datecreatedByApplication	2008-01-02	
Environment Software SwName	Macromedia Fireworks MX	maybe
Environment software swVersion	6.0	
relationship RelationshipType	structural	yes
relationship relationshipSubType	hasSibling	
relatedObjectIdentification relatedObjectIdentifierType	URI	
relatedObjectIdentification relatedObjectIdentifierValue	http://lcweb2.loc.gov/natlib/ihas/service/encyclopedia/200033287/0002v.jpg	
relatedObjectIdentification relatedObjectIdentifierSequence	2	

Example 2: 52nd Street, New York, N.Y., ca. 1948

We will only fill in a template for one of the files. A complete example would include a record for each file, showing the relationship with the relationship element. There is a relationship to the JPEG file, which is a derivative of this master file.

Data Element (Field)	**Data Value (Content)**	**Controlled Vocabulary (Yes / No)** (please specify)
objectIdentifier objectIdentifierType	URI	yes
objectIdentifier objectIdentifierValue	http://lcweb2.loc.gov/natlib/ihas/warehouse/gottlieb/02771/ver01/0001.tif	
objectCharacteristics size	60158210	
objectCharacteristics format formatDesignation formatName	image/tiff	yes
objectCharacteristics format formatDesignation formatVersion		
creatingApplication creatingApplicationName	Image Alchemy PS	
creatingApplication creatingApplicationVersion	v1.11	
creatingApplication datecreatedByApplication	20030215	
Environment Software swName	Firefox	maybe
Environment software swVersion	5.0	
relationship RelationshipType	derivation	yes
relationship relationshipSubType	source of	
relatedObjectIdentification relatedObjectIdentifierType	URI	
relatedObjectIdentification relatedObjectIdentifierValue	http://lcweb2.loc.gov/natlib/ihas/service/gottlieb/02771/ver01/0001v.jpg	
RelatedObjectIdentification relatedObjectIdentifierSequence		

Exercise for Session 4: XML syntax
Answer sheet

Example 1: America's pinch hit march (MODS)

```
<?xml version="1.0" encoding="UTF-8"?>
<mods:mods version="3.2" ID="MODS1"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/mods.xsd">
  <mods:titleInfo>
     <mods:title>America's pinch hit march</mods:title>
     <mods:subTitle>the hit that ended the world's greatest war</mods:subTitle>
  </mods:titleInfo>
  <mods:name type="personal">
    <mods:namePart>Dempsey, Bertha Stanfield</mods:namePart>
    <mods:role>
      <mods:roleTerm authority="marcrelator"
type="text">Composer</mods:roleTerm>
    </mods:role>
  </mods:name>
  <mods:typeOfResource>notated music</mods:typeOfResource>
  <mods:originInfo>
    <mods:place>
      <mods:placeTerm>Joplin, Missouri</mods:placeTerm>
    </mods:place>
    <mods:publisher>McMillan Music Co.</mods:publisher>
    <mods:dateIssued>1919</mods:dateIssued>
  </mods:originInfo>
  <mods:physicalDescription>
      <mods:form authority="ihas">sheet music</mods:form>
      <mods:extent>3 p. </mods:extent>
   </mods:physicalDescription>
  <mods:subject authority="lcsh">
      <mods:topic>Baseball</mods:topic>
      <mods:genre>Songs and music</mods:genre>
  </mods:subject>
  <mods:relatedItem type="host">
      <mods:titleInfo>
         <mods:title>Baseball sheet music collection</mods:title>
      </mods:titleInfo>
   </mods:relatedItem>
   <mods:location>
      <mods:physicalLocation authority="marcorg">DLC</mods:physicalLocation>

<mods:url>http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200033287/default.html
</mods:url>
   </mods:location>
   <mods:recordInfo>
      <mods:recordContentSource>IHAS</mods:recordContentSource>
      <mods:recordChangeDate
encoding="marc">060412</mods:recordChangeDate>
      <mods:recordIdentifier
source="IHAS">loc.natlib.ihas.200033287</mods:recordIdentifier>
```

```
  </mods:recordInfo>
</mods:mods>
```

Example 1: America's pinch hit march (DC)

```
<?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/">

   <dc:title xml:lang="en "> America's pinch hit march : the hit that ended the
worlds greatest war </dc:title>
   <dc:creator> Bertha Stanfield Dempsey</dc:creator>
   <dc:type> http://purl.org/dc/dcmitype/Image</dc:type>
   <dc:publisher>The McMillan Music Co.</dc:publisher>
   <dc:date>1919</dc:date>
   <dc:format>3 p.</dc:format>
  <dc:subject xsitype="dcterms:LCSH"> Baseball—Songs and music </dc:subject>
   <dc:relation> Baseball sheet music collection</dc:relation>

</metadata>
```

Example 2: 52nd Street, New York, N.Y., ca. 1948 (MODS)

```
<?xml version="1.0" encoding="UTF-8"?>
<mods:mods version="3.2" ID="MODS1"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/mods.xsd">
  <mods:titleInfo>
    <mods:title>52nd Street, New York, N.Y., ca. 1948</mods:title>
  </mods:titleInfo>
  <mods:name>
    <mods:namePart> Gottlieb, William P. </mods:namePart>
    <mods:namePart type="date">1917-</mods:namePart>
    <mods:role>
       <mods:roleTerm type="text"
authority="marcrelator">photographer</mods:roleTerm>
    </mods:role>
  </mods:name>
  <mods:typeOfResource>still image</mods:typeOfResource>
  <mods:originInfo>
    <mods:place>
       <mods:placeTerm>New York</placeTerm>
    </mods:place>
    <mods:dateCreated>1948</mods:dateCreated>
  <mods:physicalDescription>
    <mods:form authority="gmd">graphic</mods:form>
  </mods:physicalDescription>
  <mods:accessCondition> Original negative not served.</mods:accessCondition>
  <mods:subject authority="lcsh">
```

```
      <mods:geographic>
New York (N.Y.)</mods:geographic>
    </mods:subject>
   <mods:location>
      <mods:physicalLocation authority="marcorg">DLC</mods:physicalLocation>
   </mods:location>
   <mods:recordInfo>
       <mods:recordContentSource>DLC</mods:recordContentSource>
       <mods:recordChangeDate
encoding="marc">2007030616</mods:recordChangeDate>
       <mods:recordIdentifier source="DLC">got99000277</mods:recordIdentifier>
   </mods:recordInfo>
</mods:mods>
```

Example 2: 52nd Street, New York, N.Y., ca. 1948 (DC)

```
<?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/">

    <dc:title xml:lang="en "> 52nd Street, New York, N.Y., ca. 1948</dc:title>
    <dc:creator> Gottlieb, William P., 1917-</dc:creator>
    <dc:type> http://purl.org/dc/dcmitype/Image </dc:type>
    <dc:date>1948</dc:date>
    <dc:format>image/jpeg</dc:format>
   <dc:subject>  </dc:subject>
<dc:coverage>New York (N.Y.) </dc:coverage>
    <dc:rights Original negative not served.</dc:rights>
</metadata>
```

Note that dc:format could use image/tif or image/jpeg (or nothing) depending upon whether the object of cataloging is considered the intellectual object or the file itself.

Evaluation Form

Metadata Standards and Applications

Your evaluation of this workshop is very important to the future development of this course and other similar courses. Your honest, candid answers to the following questions will assist us in providing quality programs.

Please rate the following aspects of today's workshop by checking the box that best reflects your evaluation:

---

1. The overall content of the workshop:

	5	4	3	2	1	
a. was extremely valuable						was of little value
b. provided enough detail						was too general
c. was current & relevant						was outdated
d. was cohesive & logical						was fragmented/difficult to follow
e. was appropriate to my needs						was not at all appropriate
f. met its stated objectives						did not meet objectives

2. Presenter 1:

	5	4	3	2	1	
a. was knowledgeable						was unsure of the material
b. had good presentation skills						had poor presentation skills
c. encouraged participation						discouraged participation
d. addressed my level of understanding						did not consider my level
e. answered questions directly						did not answer questions
f. was prepared						was not prepared
g. understood the audience dynamics						ignored audience dynamics

3. Presenter 2:

	5	4	3	2	1	
a. was knowledgeable						was unsure of the material
b. had good presentation skills						had poor presentation skills
c. encouraged participation						discouraged participation
d. addressed my level of understanding						did not consider my level
e. answered questions directly						did not answer questions
f. was prepared						was not prepared
g. understood the audience dynamics						ignored audience dynamics

4. The handouts:

	5	4	3	2	1	
a. are excellent						are poor
b. followed course content						are disjointed/out of sequence
c. are valuable for future reference						are of no value

5. The PowerPoint slides:

	5	4	3	2	1	
a. were clear and easy to read						were hard to read
b. were well organized						were poorly organized
c. illustrated concepts clearly						were confusing
d. covered an appropriate amount of information						contained too much or not enough information
e. were visually effective						were not effective
f. were enhanced by and supported the presenter's remarks						were poorly related to the presenter's remarks

Please give the following information about yourself:

6. Your level of knowledge in the subject of this workshop before today: expert 5 4 3 2 1 novice
7. Your level of experience in the subject of this workshop before today: very experienced 5 4 3 2 1 beginner
8. Other comments:

Comments on specific sessions:

THANK YOU!

Note to workshop organizer: Please send copies of all completed evaluation forms to:

CLW Editorial Team Chair, The Library of Congress, Cooperative Cataloging Team, Washington, DC 20540-4373 USA