2. Descriptive Metadata Standards, Bibliographic Relationships and Metadata Models

METADATA STANDARDS AND APPLICATIONS

Objectives for Session 2

- Understand the categories of descriptive metadata standards
- Learn about various descriptive metadata standards and the communities that developed and use them
- Learn about some relationship models used in descriptive metadata standards

Descriptive Metadata

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

Metadata Standards

◆ Data structure standards (metadata schemas/formats: DC, MODS)
- Set of semantic properties, in this context used to describe resource

◆ Data exchange/syntax standards (MARC 21 (ISO 2709), DC/XML, DC/RDF)
- The structural wrapping around the semantics
- Essential for moving information around

Metadata Standards, cont.

- Data content standards (rules: AACR2R/RDA, CCO)
  - Most metadata schemas not tied to single content standard
- Data value standards (values/controlled vocabularies: LCNAF, LCSH, MeSH, AAT)
- Relationship models

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/)

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)

What is MARC 21?

◆ A syntax defined by an international standard, developed in the late 60s

◆ Two syntax expressions:
- – Classic MARC (MARC 2709)
- – MARCXML

◆ A data element set defined by content designations and semantics

◆ Institutions do not store "MARC 21", as it is a communications format

02158cam 2200349Ia 4500001001300000003000600013005001700019006001900036006001900055007001500074008004100089400020001300200015001500430021001650490009001862450119001952460025003142600065003395380030004045060038004345360153004725200764006255050094013895000086014836000049015696500040016186510039016586510023016977000026017208560050017469940012017960118051017.0 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@

MARC 21 Scope

◆ Bibliographic Data

– books, serials, computer files, maps, music, visual materials, mixed material

◆ Holdings Data

– physical holdings, digital access, location, publication history

◆ Authority Data

– names, titles, name/title combinations, subjects, series

◆ Classification Data

– classification numbers, associated captions, hierarchies

◆ Community Information

– events, programs, services, people, organizations

MARC 21 Structure

◆ The structure of MARC records is an implementation of:

– Information Interchange Format (ANSI Z39.2) and Format for Information Exchange (ISO 2709)

– Content designation (codes and conventions) as defined in the MARC 21 formats

◆ The content of most data elements is defined by standards outside the formats, for example AACR, LCSH, NLM Classification

◆ The content of other data elements (e.g., coded data), is defined in the MARC 21 formats

"The MARC21 Formats: Background and Principles"

MARC 21 in XML – MARCXML

- ◆ Established standard MARC 21 in an XML structure
  - – Takes advantage of freely available XML tools
- ◆ Allows for interoperability with different XML metadata schemas
- ◆ Provides continuity with current data
  - – XML exact equivalent of MARC (2709) record
  - – Lossless/roundtrip conversion to/from MARC 21 record
  - – Presentation using XML stylesheets
- ◆ http://ww.loc.gov/standards/marcxml

Traditional Tagged MARC 21 Record Display

000	01166cam 2200385 a 450
001	5054896
005	20041014084104.0
008	040213s2004 ilu b 001 0 eng
010	__ |a 2004003428
020	__ |a 0838908829
035	__ |a NYJH04-B3414
035	__ |a 5054896
040	__ |a DLC |c DLC |d DLC
050	00 |a Z666.5 |b .M48 2004
082	00 |a 025.3 |2 22
090	__ |a Z666.5 |b .M48 2004
245	00 |a Metadata in practice / |c Diane I. Hillmann, editor, Elaine L. Westbrooks, editor.
260	__ |a Chicago : |b American Library Association, |c 2004.
263	__ |a 0406
300	__ |a xvii, 285 p. : |b ill. ; |c 23 cm.
504	__ |a Includes bibliographical references and index.
650	_0 |a Information organization.
650	_0 |a Metadata.
650	_0 |a Database management.
700	1_ |a Hillmann, Diane I. |q (Diane Ileana), |d 1948-
700	1_ |a Westbrooks, Elaine L.
710	2_ |a American Library Association.

Dublin Core: Simple

◆ Simple to use

◆ All elements are optional/repeatable

◆ No order of elements prescribed

◆ Interdisciplinary/International

◆ Promotes semantic interoperability

◆ Controlled vocabulary <u>values</u> may be expressed, but not the sources of the values

Dublin Core Elements

◆ Fifteen elements in Simple DC

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

"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

DC Structure

◆ Property/element refinements are used at the element level in DC/XML
- – Relationships between properties and sub-properties explicit in the formal representation
- – Does not use XML "nesting" to express those relationships

◆ Encoding schemes (Syntax & Vocabulary)
- – SES: Essentially a datatype that communicates the format or structure of a string
- – VES: Includes values from an identified controlled vocabulary or list

Advantages: Dublin Core

◆ International and cross-domain

◆ Developed via an open review process

◆ Increased efficiency of the discovery/retrieval of digital objects

◆ Rich element set (qualified DC) provides a framework of elements which will aid the management of information

◆ Ease of mapping to other metadata standards promotes collaboration of cultural/educational information

Uses of Dublin Core

◆ Minimal standard for OAI-PMH

◆ Core element set in some other schemas

◆ Switching vocabulary for more complex schemas

Ex.: Simple Dublin Core

```
<metadata>
   <dc:title>Metadata in practice.</dc:title>
   <dc:contributor>Hillmann, Diane I.</dc:contributor>
   <dc:contributor>Westbrooks, Elaine L.</dc:contributor>
   <dc:subject>Information organization</dc:subject>
   <dc:subject>Metadata</dc:subject>
   <dc:subject>Database management</dc:subject>
   <dc:subject>Z666.5.M48 2004</dc:subject>
   <dc:subject>025.3</dc:subject>
   <dc:date>2004</dc:date>
   <dc:format>285 p.</dc:format>
   <dc:type>Text</dc:type>
   <dc:identifier>ISBN:0838908829</dc:identifier>
   <dc:language>en</dc:language>
   <dc:publisher>ALA Editions</dc:publisher>
</metadata>
```

Ex.: Qualified Dublin Core

```xml
<metadata>
    <dc:title xml:lang="en">Metadata in practice.</dc:title>
    <dc:contributor>Hillmann, Diane I.</dc:contributor>
    <dc:contributor>Westbrooks, Elaine L.</dc:contributor>
    <dc:subject xsitype="LCSH">Information organization</dc:subject>
    <dc:subject xsitype="LCSH">Metadata</dc:subject>
    <dc:subject xsitype="LCSH">Database management</dc:subject>
    <dc:subject xsitype="LCC">Z666.5.M48 2004</dc:subject>
    <dc:subject xsitype="DDC">025.3</dc:subject>
    <dc:date xsitype="W3CDTF">2004</dc:date>
    <dcterms:extent>285 p.</dcterms:extent>
    <dc:type xsitype="DCMIType">Text</dc:type>
    <dc:identifier xsitype="URI">ISBN:0838908829</dc:identifier>
    <dc:language xsitype="RFC3066">en</dc:language>
    <dc:publisher>ALA Editions</dc:publisher>
    <dcterms:audience>Librarians</dcterms:audience>
</metadata>
```

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)

A selection of DC projects

◆ National Science Digital Library
http://nsdl.org/
  – Aggregates a wide variety of source collections using Dublin Core
◆ Kentuckiana Digital Library
http://kdl.kyvl.org/
  – For item level metadata, on DLXS software
◆ Gathering the Jewels http://www.gtj.org.uk/
  – Website for Welsh cultural history using DC standards
◆ MusicBrainz http://musicbrainz.org/
  – User-maintained community music recording database; extension of DC

MODS

◆ MODS: Metadata Object Description Schema

◆ An XML descriptive metadata standard

◆ Derivative of MARC 21

  – Uses language based tags

  – Contains a subset of MARC elements

  – Repackages elements to eliminate some redundancies

  – Uses same "flat" record assumptions as MARC does; thus is not "FRBR-aware"

◆ Does not assume the use of any specific rules for description

◆ Element set is applicable to digital resources

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

Advantages of MODS

- ◆ Element set is largely compatible with existing MARC descriptions in large library databases

- ◆ Element set is richer than Simple Dublin Core but simpler than full MARC

- ◆ Hierarchy allows for rich description, especially of complex digital objects

- ◆ Rich description works well with hierarchical METS objects

Uses of MODS

- Extension schema to METS
  - Rich description works well with hierarchical METS objects
- As a specified XML format for SRU
- As a core element set between MARC and non-MARC XML descriptions
- For original resource description in XML syntax that is simpler than full MARC
- As an additional format when exposing information using OAI

MODS expressed in XML

```xml
<titleInfo>
    <title>Metadata in practice /</title>
</titleInfo>
<name type="personal">
    <namePart type="family">Hillmann,</namePart>
    <namePart type="given">Diane I. (Diane Ileana),</namePart>
    <namePart type="date">1948-</namePart>
    <role>
        <roleTerm type="text">editor</roleTerm>
    </role>
</name>
<name type="personal">
    <namePart type="family">Westbrooks,</namePart>
    <namePart type="given">Elaine L.</namePart>
    <role>
        <roleTerm type="text">editor</roleTerm>
    </role>
</name>
```

```xml
<typeOfResource>text</typeOfResource>
<genre authority="marc">book</genre>
<originInfo>
    <place>
        <placeTerm authority="marccountry" type="code">ilu</placeTerm>
    </place>
    <place>
        <placeTerm type="text">Chicago</placeTerm>
    </place>
    <publisher>ALA Editions</publisher>
    <dateIssued>2004</dateIssued>
    <issuance>monographic</issuance>
</originInfo>
<language>
    <languageTerm authority="iso639-2b"
type="code">eng</languageTerm>
</language>
```

```xml
<physicalDescription>
    <form authority="marcform">print</form>
    <extent>viii, 285 p. : ill. ; 23 cm.</extent>
  </physicalDescription>
  <note type="statement of responsibility">Diane I. Hillmann, editor, Elaine L.
Westbrooks, editor.</note>
  <note> Includes bibliographical references and index.</note>
  <subject authority="lcsh"><topic>Information organization</topic></subject>
  <subject authority="lcsh"><topic>Metadata</topic></subject>
  <subject authority="lcsh"><topic>Database management</topic></subject>
  <classification authority="lcc">Z666.5.M48 2004</classification>
  <classification edition="22" authority="ddc">025.3</classification>
  <recordInfo>
    <recordContentSource>DLC</recordContentSource>
    <recordCreationDate encoding="marc">20041014</recordCreationDate>
    <recordChangeDate
encoding="iso8601">20050406144503.0</recordChangeDate>
    <recordIdentifier>2004003428</recordIdentifier>
  </recordInfo>
```

Status of MODS

- Open listserv collaboration of possible implementers, LC coordinated (1st half 2002)

- First comment and use period: 2002

- Now in MODS version 3.3

- Endorsed as METS extension schema for descMD

- MODS Editorial Committee formed Fall 2008

A selection of MODS projects

- ◆ LC uses of MODS
  - – LC web archives http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html
  - – Digital library METS projects
- ◆ University of Chicago Library
  - – Chopin early editions
  - – Finding aid discovery
- ◆ Digital Library Federation Aquifer Initiative
- ◆ National Library of Australia
  - – MusicAustralia: MODS as exchange format between National Library of Australia and ScreenSoundAustralia
  - – Australian national bibliographic database metadata project

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

IEEE-LOM

◆ Nine top level categories
- General
- Life Cycle
- Meta-Metadata
- Technical
- Educational
- Rights
- Relation
- Annotation
- Classification

http://www.cancore.ca/en/guidelines.html

Advantages of IEEE-LOM

◆ Built on an explicit data model

  – Specifies which aspects of a learning object should be described

◆ International community contributes to standard

  – Education and learning sector in Europe is particularly invested in using the standard; is required in certain circumstances

◆ Applicable to a broad array of learning objects

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)

Ex.: CanCore

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

Note name & URL

Note name & URL

Status of IEEE-LOM

◆ IEEE Standard for Learning Object Metadata

- – IEEE Std 1484.12.1-2002

- – Approved 13 June 2002

- – Available for purchase from IEEE

◆ AKA Learning Resource Meta-data Specification

- – Version 1.3 (final)

- – Requires registration to download

- – http://www.imsglobal.org/metadata/

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

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.

Bibliographic Elements in ONIX

… and more!

◆ Title

◆ Author

◆ ISBN

◆ Price, availability

◆ Blurb, reviews, extracts

◆ BISAC Subject Codes

◆ Territorial rights

◆ Links to websites and book cover images

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?

Uses of ONIX

◆ ONIX in XML uses an ONIX DTD (Document Type Definition)

◆ Transmitted via email attachment or ftp

◆ Used by LC and other libraries as starting point for library descriptive cataloging

◆ New business opportunities for catalogers: Technical Service Providers will implement ONIX in companies (publishers) that lack IT expertise

Ex.: ONIX

\<Title\>
\<TitleType\>01**\</TitleType\>**
\<TitleText textcase = "02"\>British English, A to Zed**\</TitleText\>**
\</Title\>
\<Contributor\>
\<SequenceNumber\>1**\</SequenceNumber\>**
\<ContributorRole\>A01**\</ContributorRole\>**
\<PersonNameInverted\>Schur, Norman
W**\</PersonNameInverted\>**
\<BiographicalNote\>A Harvard graduate in Latin and Italian
literature, Norman Schur attended the University of Rome and the
Sorbonne before returning to the United States to study law at
Harvard and Columbia Law Schools. Now retired from legal
practise, Mr Schur is a fluent speaker and writer of both British
and American English.**\</BiographicalNote\>**
 \</Contributor\>

Ex.: ONIX, cont.

<othertext>
<d102>01**</d102>**
<d104>BRITISH ENGLISH, A TO ZED is the thoroughly updated, revised, and expanded third edition of Norman Schur's highly acclaimed transatlantic dictionary for English speakers. First published as BRITISH SELF-TAUGHT and then as ENGLISH ENGLISH, this collection of Briticisms for Americans, and Americanisms for the British, is a scholarly yet witty lexicon, combining definitions with commentary on the most frequently used and some lesser known words and phrases. Highly readable, it's a snip of a book, and one that sorts out – through comments in American – the "Queen's English" – confounding as it may seem.**</d104>**
</othertext>
<othertext>
<d102>08**</d102>**
<d104>Norman Schur is without doubt the outstanding authority on the similarities and differences between British and American English. BRITISH ENGLISH, A TO ZED attests not only to his expertise, but also to his undiminished powers to inform, amuse and entertain. – Laurence Urdang, Editor, VERBATIM, The Language Quarterly, Spring 1988 **</d104>**
</othertext>

x.: ONIX, cont.

Main Desc.

<othertext>
<d102>01**</d102>**
<d104>BRITISH ENGLISH, A TO ZED is the thoroughly updated, revised, and expanded third edition of Norman Schur's highly acclaimed transatlantic dictionary for English speakers. First published as BRITISH SELF-TAUGHT and then as ENGLISH ENGLISH, this collection of Briticisms for Americans, and Americanisms for the British, is a scholarly yet witty lexicon, combining definitions with commentary on the most frequently used and some lesser known words and phrases. Highly readable, it's a snip of a book, and one that sorts out – through comments in A   "Queen's English" – confounding as it may seem.**</d10**
</othertex
<othertext

Review

<d102>08**</d102>**
<d104>Norman Schur is without doubt the outstanding authority on the similarities and differences between British and American English. BRITISH ENGLISH, A TO ZED attests not only to his expertise, but also to his undiminished powers to inform, amuse and entertain. – Laurence Urdang, Editor, VERBATIM, The Language Quarterly, Spring 1988 **</d104>**
</othertext>

Status of ONIX for Books

- ONIX For Books, Release 2.1 revision 02 (Feb 2005) for most
- BISAC identifies 31 data elements as best practice
- ONIX/MARC21 mappings by Library of Congress, OCLC
- ONIX/UNIMARC mapping by British Library
  - http://www.editeur.org/onixmarc.html

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

What is 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 for discovery, exchange and use of data

  – Single-level description at collection level
  – Multilevel description from collection through file and item levels

◆ Developed and maintained by Society of American Archivists

  – LC hosts the website: http://www.loc.gov/ead/

Advantages of EAD

◆ Documents explicitly 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

Uses of EAD

- ◆ Based on the needs of the archival community
  - – Hierarchical structure difficult to map to non-hierarchical schemas
  - – Lower levels assume some inheritance of information, can't be mapped separately
- ◆ Good at describing blocks of information, poor at providing granular information
- ◆ Some uptake by museum community
- ◆ Provides standard method for re-using existing printed Finding Aids in a digital environment

EAD Example

<eadheader audience="internal" countryencoding="iso3166-1" dateencoding="iso8601" langencoding="iso639-2b" relatedencoding="DC" repositoryencoding="iso15511" scriptencoding="iso15924">
   <eadid countrycode="us" identifier="bachrach_lf" mainagencycode="NSyU">bachrach_lf</eadid> <filedesc>
     <titlestmt>
        <titleproper encodinganalog="Title">Louis Fabian Bachrach Papers</titleproper>
        <subtitle>An inventory of his papers at Blank University</subtitle>
        <author encodinganalog="Creator">Mary Smith</author>
     </titlestmt>

-         <publicationstmt>

EAD Example, cont.

`<publisher encodinganalog="Publisher">Blank University</publisher>`

`<date encodinganalog="Date" normal="1981">1981</date>`

`</publicationstmt>`

`</filedesc>`

`<profiledesc>`

`<creation>John Jones`

`<date normal="2006-09-13">13 Sep 2006</date>`

`</creation>`

`<language>`

`<language encodinganalog="Language" langcode="eng">English </language>`

`</langusage>`

`</profiledesc>`

`</eadheader>`

Status of EAD

◆ EAD 2002 (current)

◆ EAD Roundtable of SAA promotes implementation and use of EAD.DTD

◆ Design Principles for enhancements to EAD enable to grow rationally

– http://www.loc.gov/ead/eaddesgn.html

A Selection of EAD Projects

◆ Archives Hub [UK] http://www.archiveshub.ac.uk/

◆ Library of Congress Finding Aids Project http://www.loc.gov/rr/ead/

◆ Cartoon Research Library, Ohio State University http://cartoons.osu.edu/

◆ Virginia Heritage http://www.lib.virginian.edu/vhp/

VRA Core

◆ Maintained by the Visual Resources Association

◆ 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

VRA: 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

Advantages of VRA Core

- ◆ Allows description of original and digital object

- ◆ Level of granularity supports specific discipline

- ◆ New content rules have been developed: Cataloging Cultural Objects (CCO)

  - – Extensive community developing around the standards: http://vraweb.org/ccoweb/cco/index.html

Uses of VRA Core

- Standardized description of art works
  - Includes description of analog objects that have been digitized
- Some acceptance by museums, art collections
- CCO standard based on print publication model
  - "Selections" and examples on web page, but full standard not available there

Ex.: 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>
```

```
<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>
```

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>
```

```
<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>
```

Status of VRA Core

- ◆ Version 4.0 (9 April 2007)
  - – http://www.vraweb.org/projects/vracore4/
- ◆ Endorsed by METS as an official extension schema for images of cultural heritage resources
- ◆ Unrestricted version
  - – Specifies the basic structure of the schema
- ◆ Restricted version
  - – Adds controlled type lists and date formats

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/

PB Core

- ◆ 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

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

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

PB Core

```
<PBCoreDescriptionDocument xsi:schemaLocation=
"http://www.pbcore.org/PBCore/PBCoreNamespace.html
http://www.pbcore.org/PBCore/PBCoreSchema.xsd">
<pbcoreIdentifier>
<identifier>8sMPyUcqFUTvVHzq1bNB</identifier>
<identifierSource>PBCore Cataloging Tool</identifierSource>
</pbcoreIdentifier>
<pbcoreTitle>
<title>Secrets of the Lost Canyon</title>
<titleType>Program</titleType>
</pbcoreTitle>
<pbcoreSubject>
<subject>
Prehistoric Native American culture; Desert Cultures; Ancestral Puebloans;
Anasazi; Fremont culture; Prehistoric art; Early Native American arts and
crafts; Prehistoric Native American arts and crafts; Desert Cultures; Desert
Gatherer arts and crafts; Range Creek Canyon, Utah
</subject>
<subjectAuthorityUsed>Library of Congress Subject
Headings</subjectAuthorityUsed>
</pbcoreSubject>
<pbcoreDescription>
```

```
<description>
When the existence of Range Creek Canyon, with its hundreds--if not
thousands--of ancient, undisturbed, Fremont Indian sites was announced in
the Summer of 2004, worldwide interest focused on a unique parcel of land
wedged in a remote corner of Utah. A Utah ranching family had defied the
pressures of encroaching modern society and [portion omitted]
</description>
<descriptionType>Abstract</descriptionType>
</pbcoreDescription>
<pbcoreGenre>
<genre>
Documentary; Educational; History; Politics; Western; Nature; Science;
Environment
</genre>
<genreAuthorityUsed>PBCore Genre Picklist</genreAuthorityUsed>
</pbcoreGenre>
<pbcoreRelation>
<relationType>relationType0</relationType>
<relationIdentifier>relationIdentifier0</relationIdentifier>
</pbcoreRelation>
```

… and so forth …

Status of PBCore

- Version 1.1 (1st quarter 2007)
- Freely available to use under Creative Commons license, with attribution
- National Center for Accessible Media (NCAM)/WGBH Educational Foundation is authority & maintenance organization
- Listserv: pbcore-users

Selection of PBCore projects

- ◆ Wisconsin Public Television (WPT) Media Library Online http://wptmedialibrary.wpt.org/

- ◆ Kentucky Educational Television (KET) http://www.ket.org/

- ◆ New Jersey Network (NJN) http://www.njn.net/

Text Encoding Initiative (TEI)

◆ Consortium of institutions and research projects

◆ Maintains and develops guidelines for the representation of texts in digital form

◆ 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 & the original source

TEI Header

- File description <fileDesc>
  - Required for TEI header
  - title, edition, extent, publication, series, notes, source
- Encoding description <encodingDesc>
- Text profile <profileDesc>
- Revision history <revisionDesc>

Advantages of using TEI

- Hardware-, software-, and application-independent
- Conventions for many key text types and features
- Applicable to many disciplines and purposes (computational linguistics, literary analysis, theological textual analysis)

Uses of TEI

- Document texts for scholars, catalogers in libraries and archives, and software to process the texts

- Equivalent of code books or manuals that accompany electronic data sets

- TEI allows scholars to analyze texts, test authenticity, compare versions

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>
    <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>
```

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

Status of TEI

- ◆ TEI P4 (2004)
  - XML-compatible
- ◆ Guidelines and resources for learning TEI available
  - http://www.tei-c.org/index.xml
- ◆ Most intensively used in the humanities

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/

Modeling metadata: why use models?

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

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)

Bibliographic relationships (pre-FRBR)

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

Bibliographic relationships in MARC/MODS

◆ MARC Linking entry fields

◆ MARC relationships by specific encoding formats
  – Authority/bibliographic/holding

◆ MODS relationships
  – relatedItem types
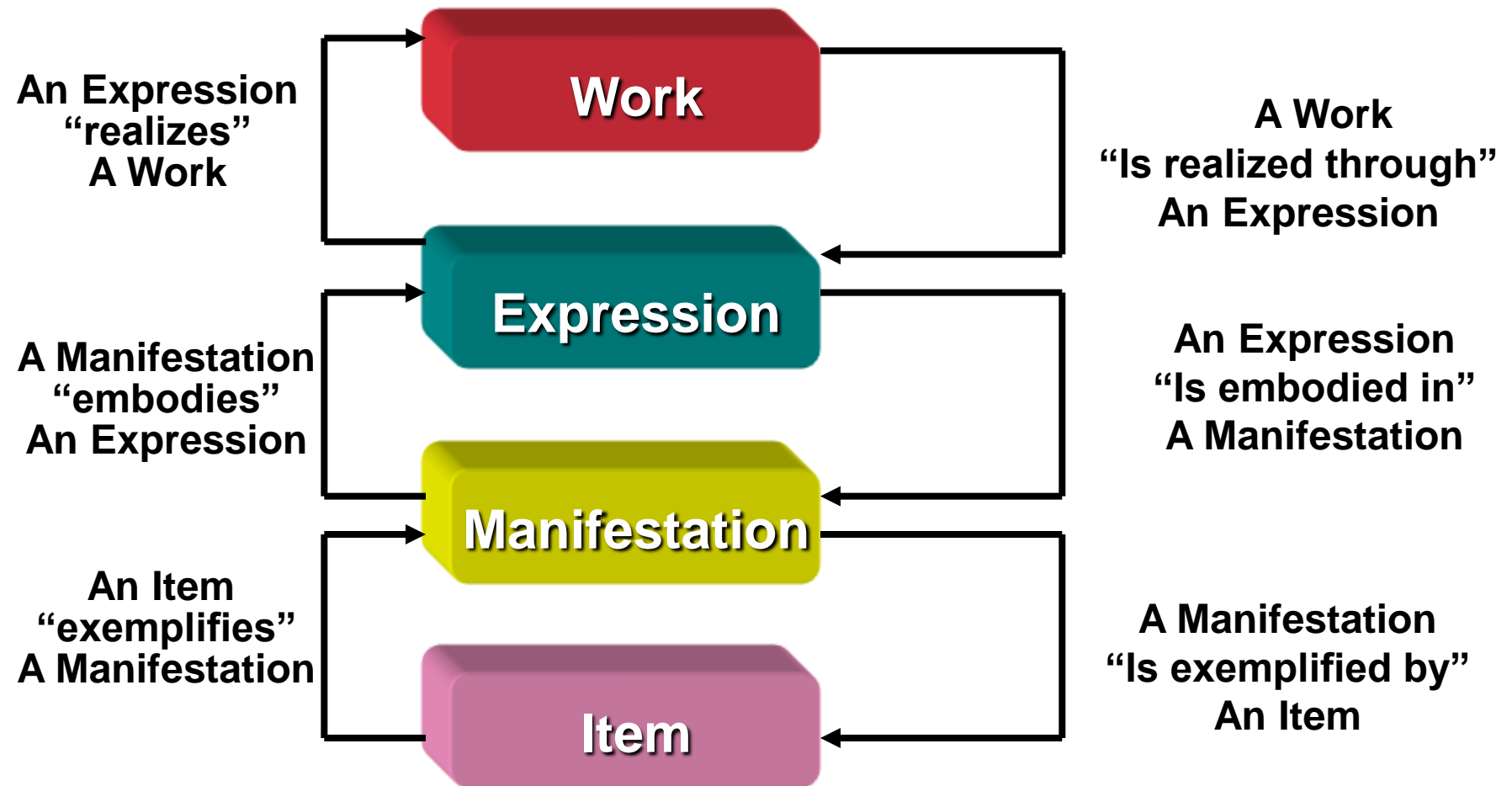 – Additional structural relationships when used in METS documents

FRBR (1996)

◆ IFLA Study Group on the Functional Requirements for Bibliographic Records

◆ Focused on the bibliographic record rather than the catalog

◆ Used an entity relationship model, rather than descriptive analysis without a structural model

◆ Broader in scope than previous studies

FRBR Entities

◆ Bibliographic entities: *works, expressions, manifestations, items*

◆ Responsible parties: *persons, corporate bodies*

◆ Subject entities: *concepts, objects, events, places*

Group 1 Entities & Their Relationships

Work

Expression

Manifestation

Item

**An Expression
"realizes"
A Work**

**A Work
"Is realized through"
An Expression**

**A Manifestation
"embodies"
An Expression**

**An Expression
"Is embodied in"
A Manifestation**

**An Item
"exemplifies"
A Manifestation**

**A Manifestation
"Is exemplified by"
An Item**

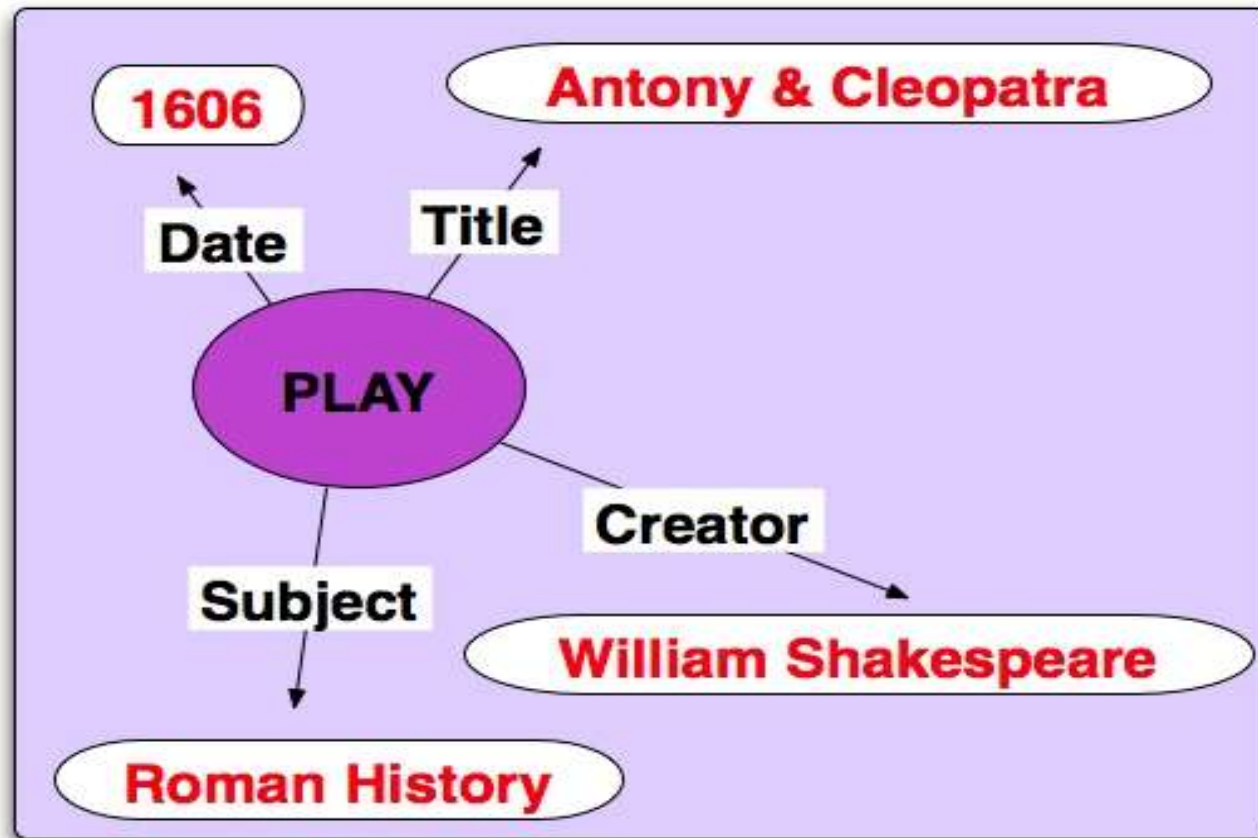*[Thanks to Sherry Vellucci for this slide.]*

DC Abstract Model

- ◆ Reaffirms the One-to-One Principle
- ◆ Defines 'statement' as the atomic level
- ◆ Distinguishes between "description" and "description set":
  - – **Description**: "One or more statements about one, and only one, resource."
  - – **Description Set**: "A set of one or more descriptions, each of which describes a single resource."
- ◆ RDA vocabularies being developed to use the DC Abstract Model

is instantiated as

record

is grouped into

description set
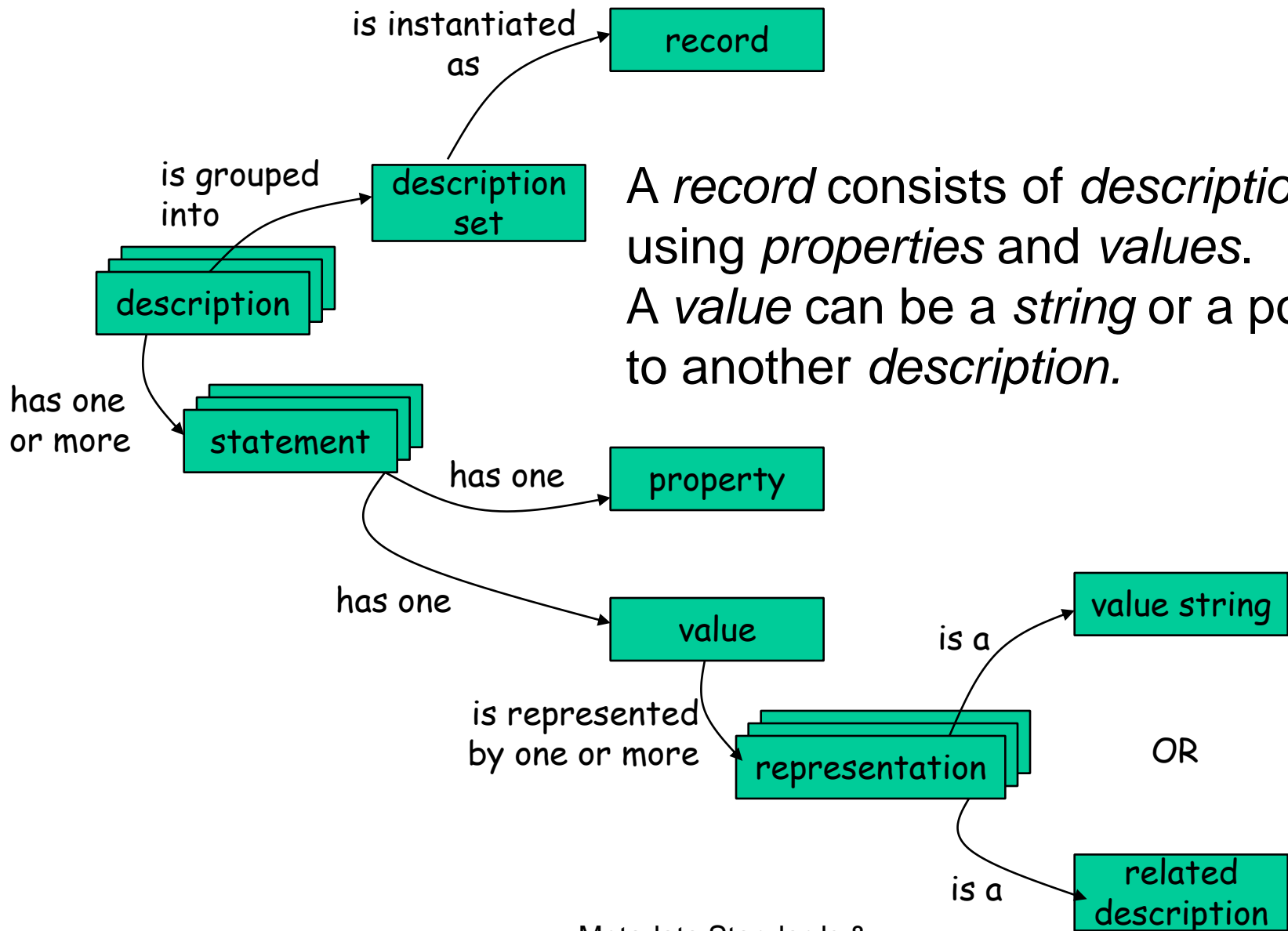
description

A *record* consists of *descriptions*, using *properties* and *values*.
A *value* can be a *string* or a pointer to another *description.*

has one or more

statement

has one

property

has one

value

is a

value string

is represented by one or more

representation

OR

is a

related description

Metadata Standards & Applications

89

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
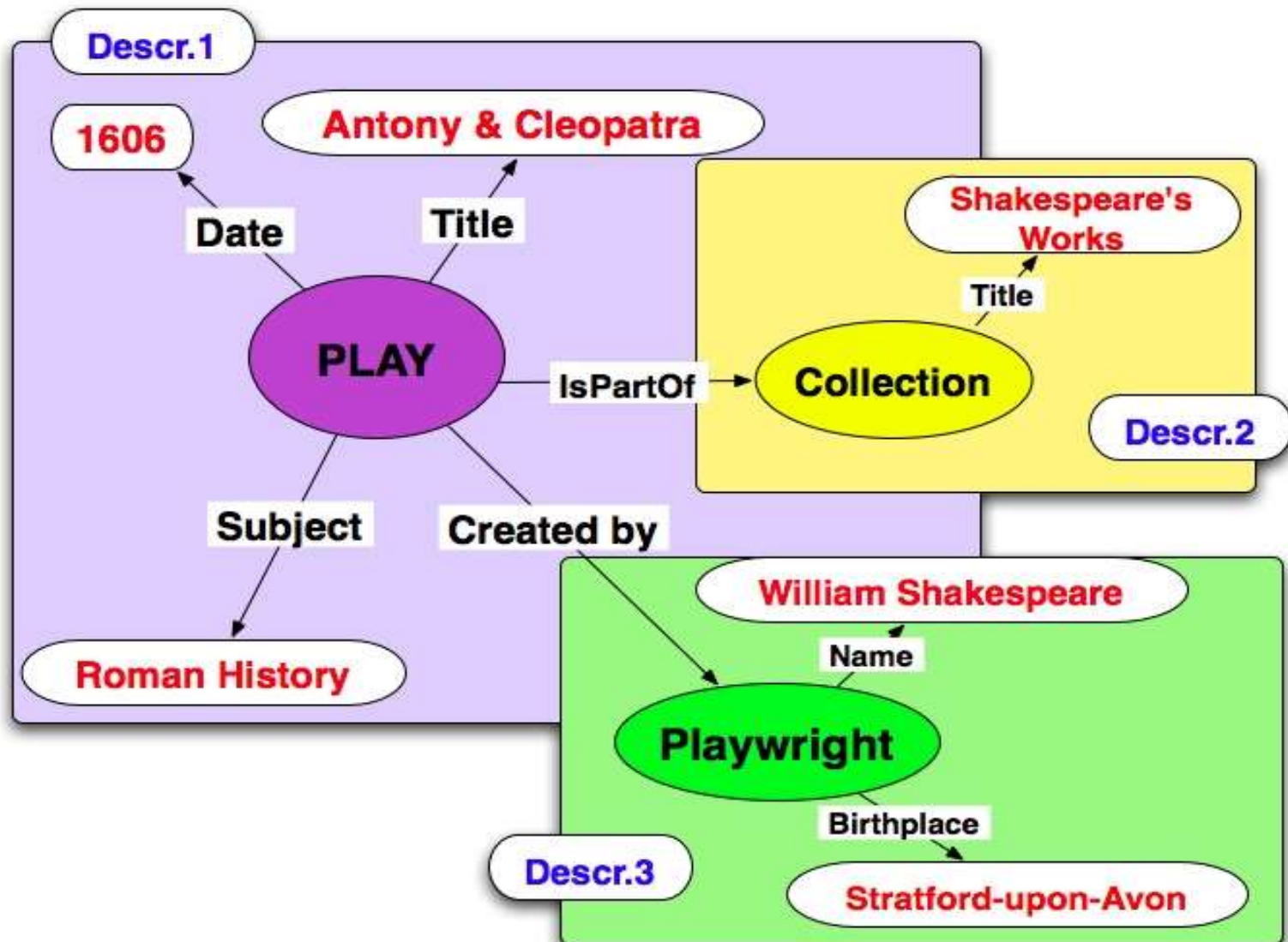
An Exercise

◆ *Each group will be given a printout of a digital object*

◆ *Create a brief metadata record based on the standard assigned to your group*

◆ *Take notes about the issues and decisions made*

◆ *Appoint a spokesperson to present the metadata record created & the issues involved (5-10 minutes)*