

## Learning Objectives

- ❖ Understand the roles of schemas and DTDs in defining tags, attributes and structures used in defining metadata standards
- ❖ List types of metadata and their roles in:
  - describing resources
  - defining rights and controlling access to resources
  - providing access to resources
  - preserving and maintaining resources

3

## Well-formed vs. Valid

- ❖ An XML document must be well-formed but doesn't have to be tied to a schema or DTD (i.e., valid).
- ❖ Well-formed XML conforms to the requirements outlined in *Metadata Basics*.
- ❖ Valid XML is well-formed, and is also validated (parsed) against a DTD or a Schema.
  - To conform to a specific application of XML

4

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="anthology.rng" type="application/xml"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<anthology>
  <poem>
    <title>The SICK ROSE</title>
    <stanza>
      <line>O Rose thou art sick.</line>
      <line>The invisible worm,</line>
      <line>That flies in the night</line>
      <line>In the howling storm:</line>
    </stanza>
    <stanza>
      <line>Has found out thy bed</line>
      <line>Of crimson joy:</line>
      <line>And his dark secret love</line>
      <line>Does thy life destroy.</line>
    </stanza>
  </poem>
  <!-- more poems go here -->
</anthology>

```

## XML Example

From "A Gentle Introduction to XML."  
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>

5

## DTD for Poems Anthology

<!ELEMENT anthology	(poem+)>
<!ELEMENT poem	(title?, stanza+)>
<!ELEMENT title	(#PCDATA) >
<!ELEMENT stanza	(line+) >
<!ELEMENT line	(#PCDATA) >

### DTD Symbols

+ = 1 or more, e.g., one or more poems go in anthology  
 ? = 0 or 1—a poem may or may not have a title  
 , = sequence, e.g., title precedes stanza  
 \* = 0 or more  
 | = or

6

## Relax NG (Compact) Schema for Poems Anthology

```

namespace rng = "http://relaxng.org/ns/structure/1.0"
datatypes xsd = "http://www.w3.org/2001/XMLSchema-datatypes"
start = anthology
anthology = element anthology {poem+}
poem = element poem {title?, stanza+}
title = element title { text }
stanza = element stanza { line+ }
line = element line { text }

```

### Relax NG Compact Symbols

+ = 1 or more, e.g., one or more poems go in anthology  
 ? = 0 or 1—a poem may or may not have a title  
 , = sequence, e.g., title precedes stanza  
 \* = 0 or more  
 | = or

7

```

<?xml version="1.0" encoding="UTF-8"?>
<grammar xmlns="http://relaxng.org/ns/structure/1.0"
  xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0"
  datatypeLibrary="http://www.w3.org/2001/XMLSchema-datatypes">
  <start>
    <element name="anthology">
      <oneOrMore>
        <element name="poem">
          <optional>
            <element name="title">
              <text />
            </element>
          </optional>
          <oneOrMore>
            <element name="stanza">
              <oneOrMore>
                <element name="line">
                  <text />
                </element>
              </oneOrMore>
            </element>
          </oneOrMore>
        </element>
      </oneOrMore>
    </element>
  </start>
</grammar>

```

## RelaxNG Schema for Anthology

8

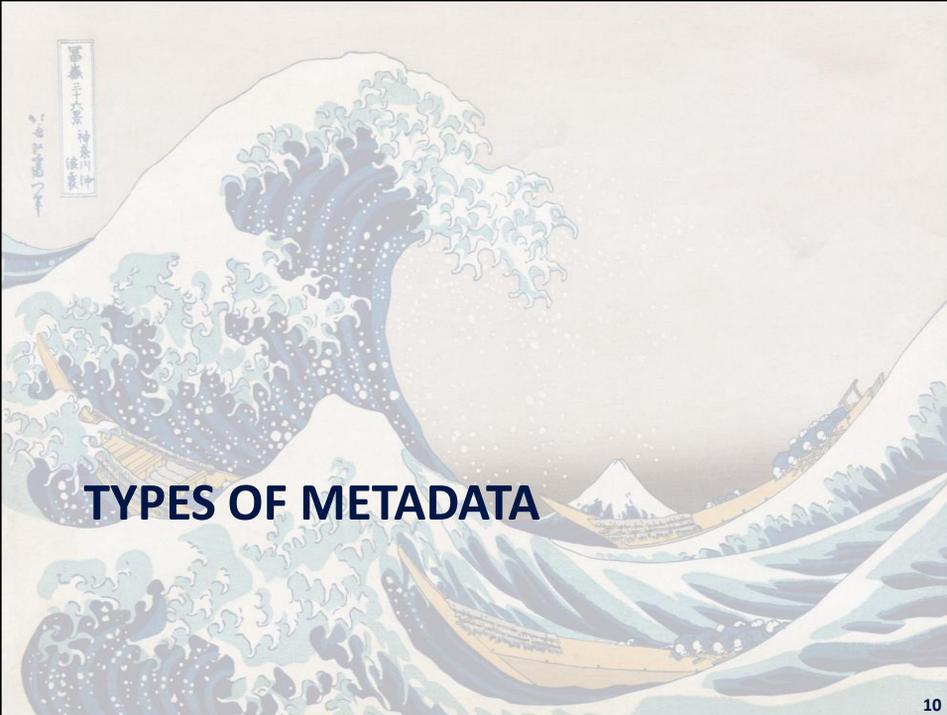
```

<?xml version="1.0" encoding="UTF-8"?>
<xs:schema elementFormDefault="qualified" attributeFormDefault="unqualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="anthology">
    <xs:annotation>
      <xs:documentation>Defines an anthology of poetry</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence maxOccurs="unbounded">
        <xs:element name="poem" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="poem">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="title" minOccurs="0" maxOccurs="1" />
        <xs:element ref="stanza" maxOccurs="unbounded" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="title" type="xs:string" />
  <xs:element name="stanza">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="line" maxOccurs="unbounded" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="line" type="xs:string" />
</xs:schema>

```

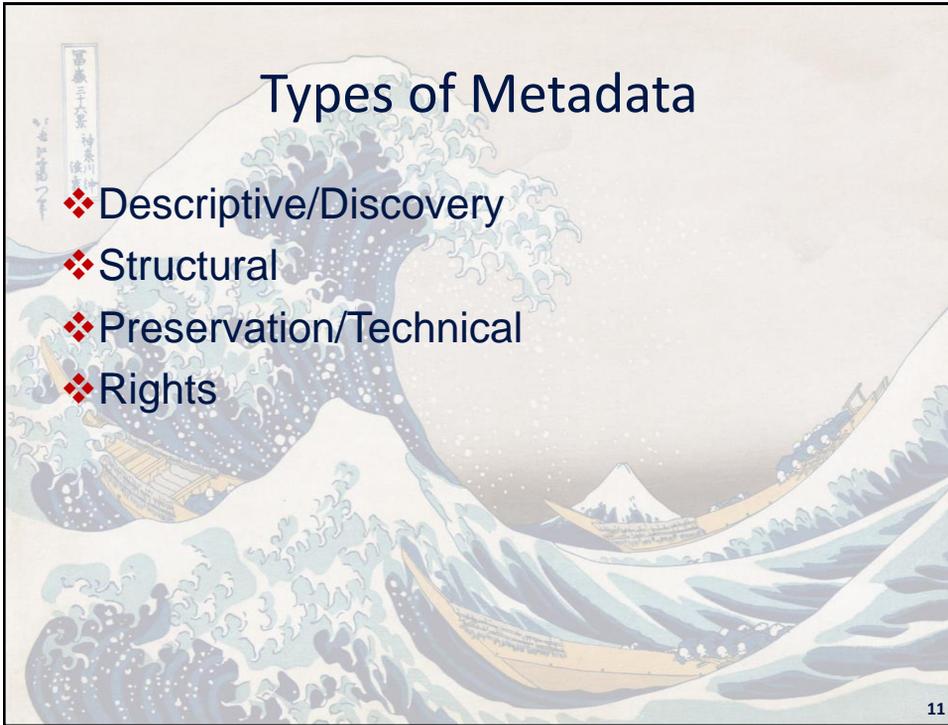
## XML Schema

9



## TYPES OF METADATA

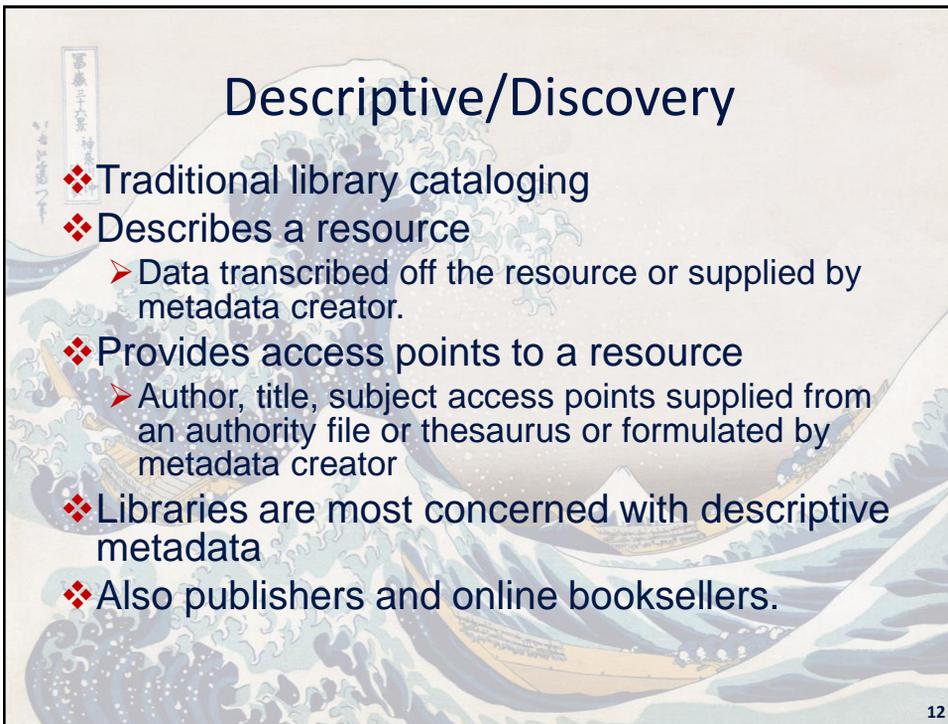
10



## Types of Metadata

- ❖ Descriptive/Discovery
- ❖ Structural
- ❖ Preservation/Technical
- ❖ Rights

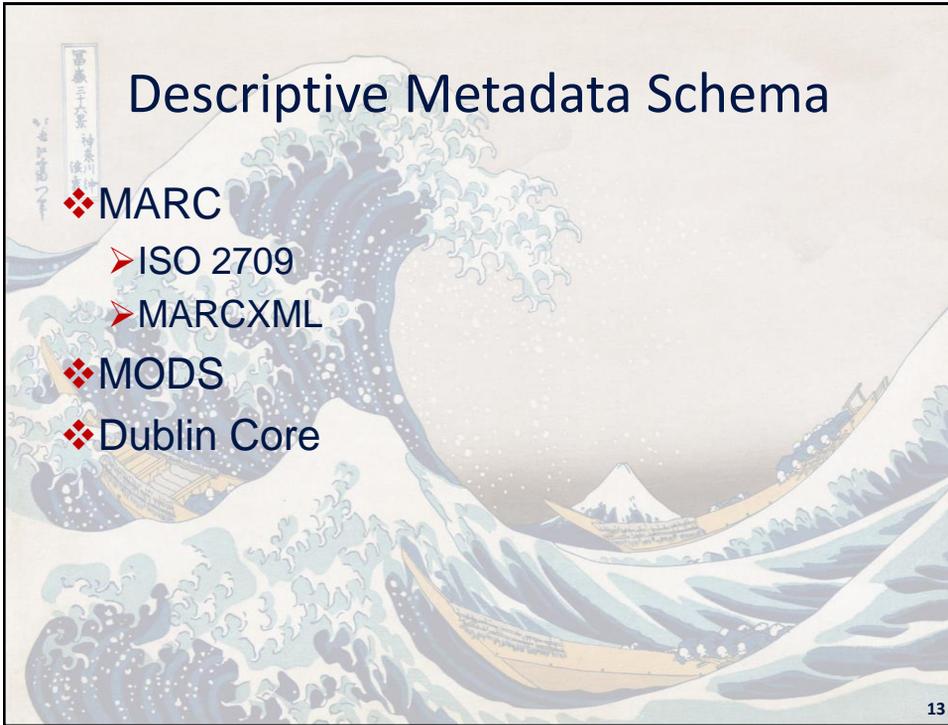
11



## Descriptive/Discovery

- ❖ Traditional library cataloging
- ❖ Describes a resource
  - Data transcribed off the resource or supplied by metadata creator.
- ❖ Provides access points to a resource
  - Author, title, subject access points supplied from an authority file or thesaurus or formulated by metadata creator
- ❖ Libraries are most concerned with descriptive metadata
- ❖ Also publishers and online booksellers.

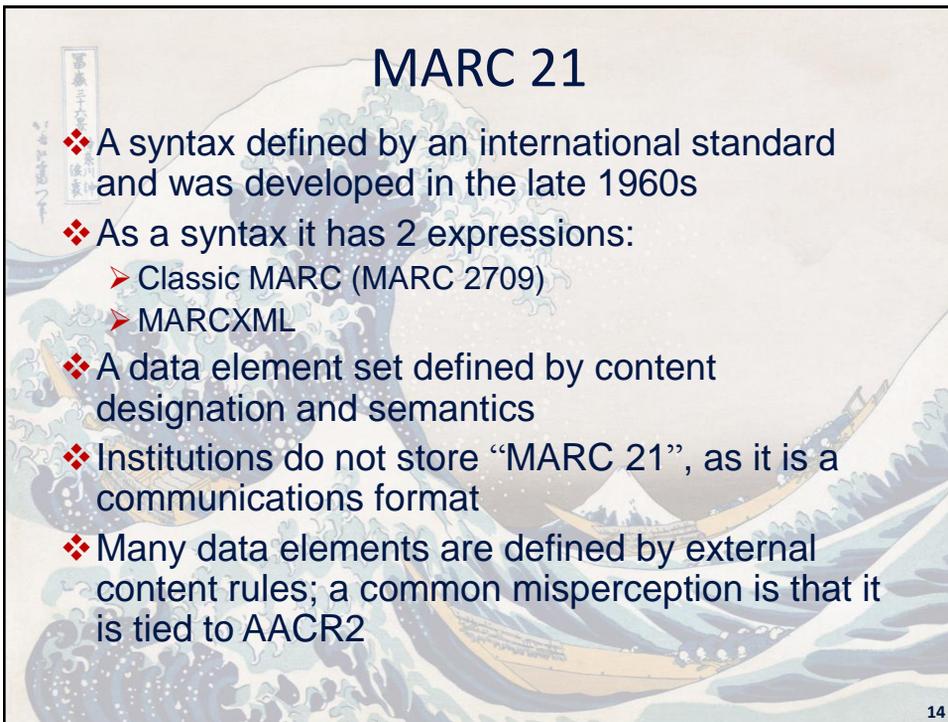
12



## Descriptive Metadata Schema

- ❖ MARC
  - ISO 2709
  - MARCXML
- ❖ MODS
- ❖ Dublin Core

13



## MARC 21

- ❖ A syntax defined by an international standard and was developed in the late 1960s
- ❖ As a syntax it has 2 expressions:
  - Classic MARC (MARC 2709)
  - MARCXML
- ❖ A data element set defined by content designation and semantics
- ❖ Institutions do not store “MARC 21”, as it is a communications format
- ❖ Many data elements are defined by external content rules; a common misperception is that it is tied to AACR2

14

## ISO 2709 MARC Record

01239cam 2200313 i 4500  
 0010007000000050017000070080041000240350021000659060045000860100017  
 00131020002400148040001800172050002000190082001600210**100002400226**  
**245010400250260003900354300003300393**50400320042650000200045860  
 0007900478600007600557600004400633650005900677650005300736650004200  
 78965000350083199100590086639178620050201142913.0790502s1979 maua  
 b 001 0 eng 9(DLC) 79013858 a7bcbccorignewd1eocipf19gy-gencatlg a  
 79013858 a039528208X :c\$8.95  
 aDLCcDLCdDLC00aPR6039.O32bZ7800a828/.9/12091 a**O'Neill, Timothy R.**  
**14aThe individuated hobbit :bJung, Tolkien, and the archetypes of Middle-**  
**Earth /cTimothy R. O'Neill. aBoston :bHoughton Mifflin,c1979. axv, 200 p.**  
**:bill. ;c22 cm. aBibliography: p. [183]-184. aIncludes index.10aTolkien, J. R.**  
 R.q(John Ronald Reuel),d1892-1973xKnowledgexPsychology.10aTolkien, J. R.  
 R.q(John Ronald Reuel),d1892-1973.tLord of the rings.10aJung, C. G.q(Carl  
 Gustav),d1875-1961. 0aPsychological fiction, EnglishxHistory and criticism.  
 0aFantasy fiction, EnglishxHistory and criticism. 0aArchetype (Psychology) in  
 literature. 0aMiddle Earth (Imaginary place) bc-  
 GenCollhPR6039.O32iZ78p0001683919AtCopy 1wBOOKS

15

## MARC (Human View)

[010](#) 79013858  
[040](#) DLC \$c DLC \$d BAKER \$d BTCTA \$d YDXCP \$d OCLCG  
[020](#) 039528208X : \$c \$8.95  
[020](#) 9780395282083  
[050](#) 00 PR6039.O32 \$b Z78  
[082](#) 00 828/.9/1209  
[049](#) TPQQ  
[100](#) 1\_ O'Neill, Timothy R.  
[245](#) 14 The individuated hobbit : \$b Jung, Tolkien, and the archetypes of  
 Middle-Earth / \$c Timothy R. O'Neill.  
[260](#) Boston : \$b Houghton Mifflin, \$c 1979.  
[300](#) xv, 200 p. : \$b ill. ; \$c 22 cm.  
[504](#) Bibliography: p. [183]-184.  
[500](#) Includes index.  
[530](#) Also issued online.

16

## 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

## MARCXML—Control Fields

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"
xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>01239cam 2200313 i 4500</leader>
    <controlfield tag="001">391786</controlfield>
    <controlfield tag="005">20050201142913.0</controlfield>
    <controlfield tag="008">790502s1979 maua
      b 001 0 eng </controlfield>
    <datafield tag="035" ind1=" " ind2=" ">
      <subfield code="9">(DLC) 79013858</subfield>
    </datafield>
```

18

## MARCXML—Selected Data Fields

```

<datafield tag="100" ind1="1" ind2=" " >
  <subfield code="a">O'Neill, Timothy R.</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="4">
  <subfield code="a">The individuated hobbit :</subfield>
  <subfield code="b">Jung, Tolkien, and the archetypes of
    Middle-Earth /
  </subfield>
  <subfield code="c">Timothy R. O'Neill.</subfield>
</datafield>
<datafield tag="260" ind1=" " ind2=" " >
  <subfield code="a">Boston :</subfield>
  <subfield code="b">Houghton Mifflin,</subfield>
  <subfield code="c">1979.</subfield>
</datafield>

```

19

## 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

20

## 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

21

## MODS

### Advantages

- ❖ 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

### Limitations

- ❖ Multiple levels of tags difficult to leverage in digital systems
- ❖ Hence, no MODS-based turnkey systems
- ❖ Most MODS-based systems custom created by individual institutions
- ❖ No round-tripability with MARC 21 or Dublin Core

22

## Uses of MODS

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

23

## Example: MODS

```
<?xml version="1.0" encoding="UTF-8"?>
<mods:modsCollection 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-3-3.xsd"
  xmlns:mods="http://www.loc.gov/mods/v3">
  <mods:mods ID="nh001tiff">
    <mods:titleInfo>
      <mods:title>Aya and Students at Yatsushiro</mods:title>
    </mods:titleInfo>
    <mods:name type="personal">
      <mods:namePart type="family">Walker,</mods:namePart>
      <mods:namePart type="given">Bill</mods:namePart>
      <mods:namePart type="date">1954-</mods:namePart>
    </mods:name>
  </mods:mods>
</mods:modsCollection>
```

24

## Example: MODS

```

<mods:originInfo>
  <mods:dateCaptured encoding="iso8601">2005-11-22
  </mods:dateCaptured>
</mods:originInfo>
<mods:physicalDescription>
  <mods:digitalOrigin>born digital</mods:digitalOrigin>
  <mods:internetMediaType>
    image/tiff
  </mods:internetMediaType>
  <mods:extent>5.5 MB</mods:extent>
  <mods:reformattingQuality>
    preservation
  </mods:reformattingQuality>
</mods:physicalDescription>
<mods:note>
  1000 x 750 pixels; 253 KB; 24-bit color
</mods:note>
<mods:abstract>
  Aya Masuda and some of her girl students at the Yatsushiro
  Fall Festival
</mods:abstract>

```

25

## Example: MODS

```

<mods:subject authority="lcsh">
  <mods:topic>Junior high school students</mods:topic>
  <mods:geographic>Japan</mods:geographic>
</mods:subject>
<mods:subject authority="lcsh">
  <mods:topic>Junior high school teachers</mods:topic>
  <mods:geographic>Japan</mods:geographic>
</mods:subject>
<mods:subject authority="lcsh">
  <mods:geographic>Yatsushiro-shi (Japan)</mods:geographic>
</mods:subject>
<mods:subject authority="lcsh">
  <mods:name>
    <mods:namePart type="family">Masuda,</mods:namePart>
    <mods:namePart type="given">Aya</mods:namePart>
    <mods:namePart type="date">1978-</mods:namePart>
  </mods:name>
</mods:subject>

```

26

## Example: MODS

```

<mods:relatedItem ID="nh001jpeg">
  <mods:physicalDescription>
    <mods:extent>1 photograph (253 KB)</mods:extent>
    <mods:internetMediaType>
      image/jpeg
    </mods:internetMediaType>
    <mods:reformattingQuality>
      access
    </mods:reformattingQuality>
    <mods:note>1000 x 750 pixels; 24-bit color</mods:note>
  </mods:physicalDescription>
</mods:relatedItem>
<mods:relatedItem ID="nh001gif">
  <mods:physicalDescription>
    <mods:extent>1 photograph (44 KB)</mods:extent>
    <mods:internetMediaType>
      image/jpeg
    </mods:internetMediaType>
    <mods:reformattingQuality>
      access
    </mods:reformattingQuality>
    <mods:note>267 x 200 pixels; 24-bit color</mods:note>
  </mods:physicalDescription>
</mods:relatedItem>
</mods:mods>
</mods:modsCollection>

```

27

## Status of MODS

- ❖ Open listserv collaboration of possible implementers, LC coordinated
- ❖ First comment and use period: 2nd half 2002
- ❖ Current version 3.5
- ❖ Companion for authority metadata (MADS) in version 2.0
- ❖ Endorsed as METS extension schema for descMD
- ❖ Many expose records as MODS in OAI
- ❖ MODS Editorial Committee

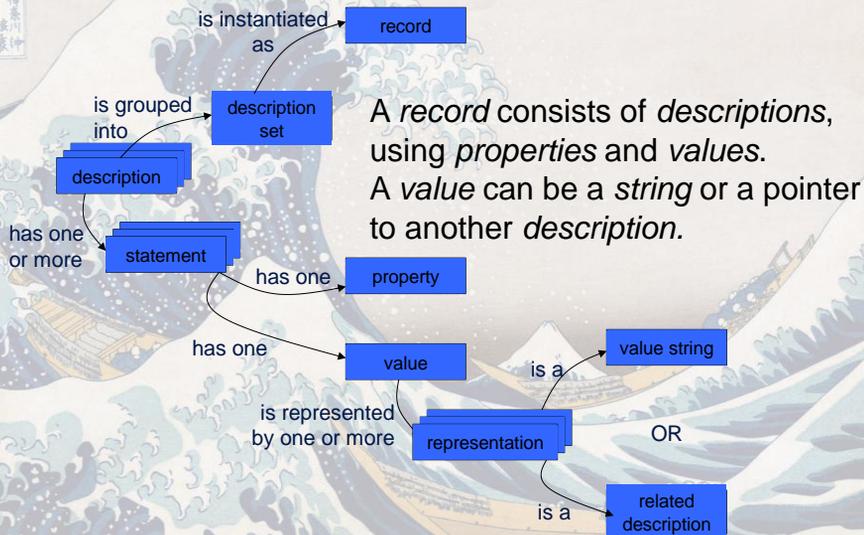
28

## Dublin Core 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

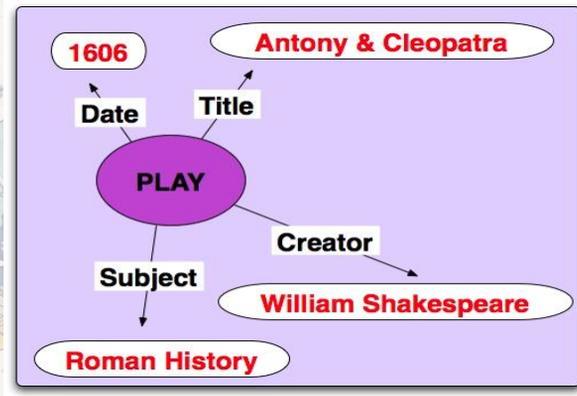
29

## DCAM



30

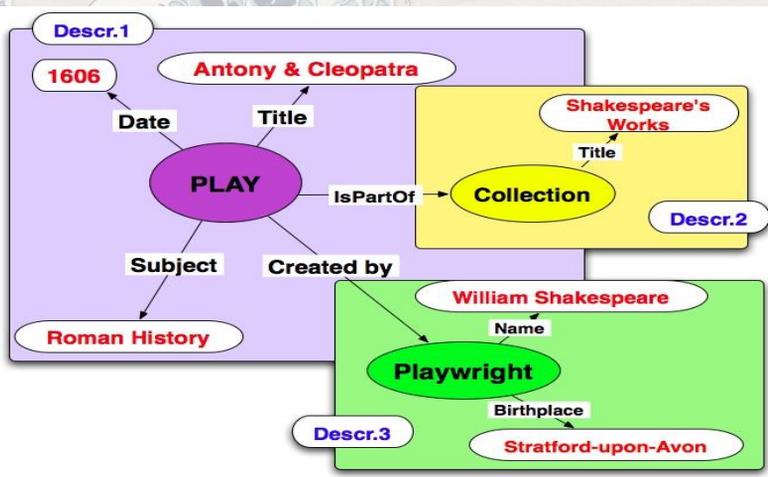
# Resource with properties



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

... related to other Resources

# Additional Properties



## Dublin Core Elements

### Fifteen elements in Simple DC

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

33

## Dublin Core “Terms”

- ❖ Includes 15 terms of the original DC Metadata Element Set, plus:
  - Properties in the terms namespace which refine the 15 DC elements
  - Vocabulary encoding schemes: DCMIType, DDC, IMT, LCC, LCSH, MESH, NLM, TGN, UDC
  - Syntax encoding schemes: Box, ISO3166, ISO639-2, ISO639-3, Period, Point, RFC1766, RFC3066, RFC4646, URI, W3CDTF
  - Classes—A class is a group containing members that have attributes, behaviors, relationships or semantics in common; a kind of category.

34

## Dublin Core Terms

- ❖ Term—the generic name for a property (i.e. element or element refinement)

[abstract](#), [accessRights](#), [accrualMethod](#), [accrualPeriodicity](#), [accrualPolicy](#), [alternative](#), [audience](#), [available](#), [bibliographicCitation](#), [conformsTo](#), [contributor](#), [coverage](#), [created](#), [creator](#), [date](#), [dateAccepted](#), [dateCopyrighted](#), [dateSubmitted](#), [description](#), [educationLevel](#), [extent](#), [format](#), [hasFormat](#), [hasPart](#), [hasVersion](#), [identifier](#), [instructionalMethod](#), [isFormatOf](#), [isPartOf](#), [isReferencedBy](#), [isReplacedBy](#), [isRequiredBy](#), [issued](#), [isVersionOf](#), [language](#), [license](#), [mediator](#), [medium](#), [modified](#), [provenance](#), [publisher](#), [references](#), [relation](#), [replaces](#), [requires](#), [rights](#), [rightsHolder](#), [source](#), [spatial](#), [subject](#), [tableOfContents](#), [temporal](#), [title](#), [type](#), [valid](#)

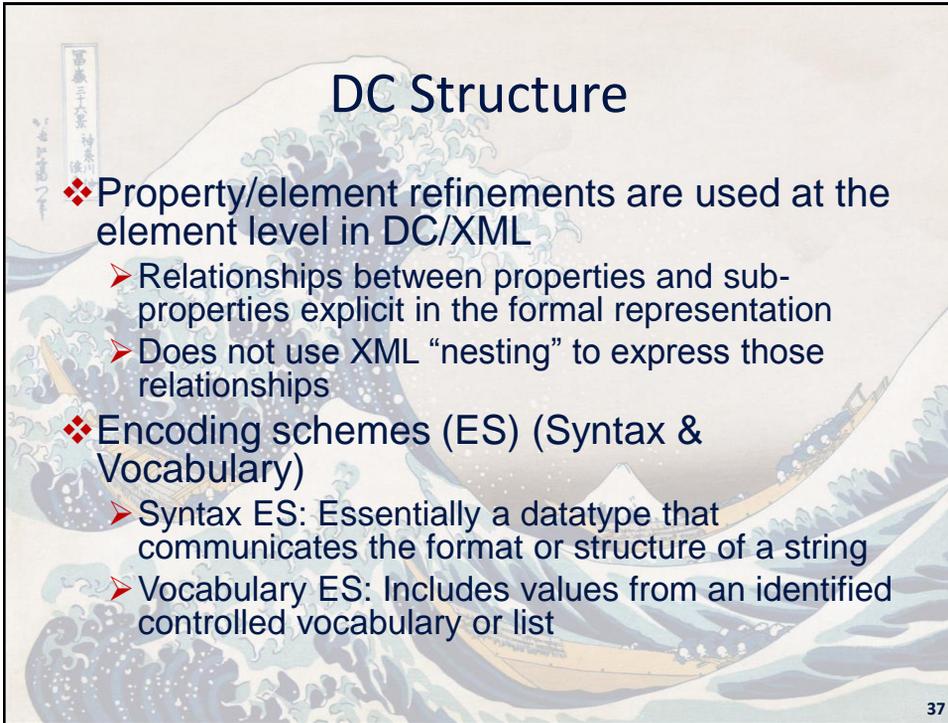
35

## Dublin Core Classes

- ❖ Class—a group containing members that have attributes, behaviors, relationships or semantics in common; a kind of category.

[Agent](#), [AgentClass](#), [BibliographicResource](#), [FileFormat](#), [Frequency](#), [Jurisdiction](#), [LicenseDocument](#), [LinguisticSystem](#), [Location](#), [LocationPeriodOrJurisdiction](#), [MediaType](#), [MediaTypeOrExtent](#), [MethodOfAccrual](#), [MethodOfInstruction](#), [PeriodOfTime](#), [PhysicalMedium](#), [PhysicalResource](#), [Policy](#), [ProvenanceStatement](#), [RightsStatement](#), [SizeOrDuration](#), [Standard](#)

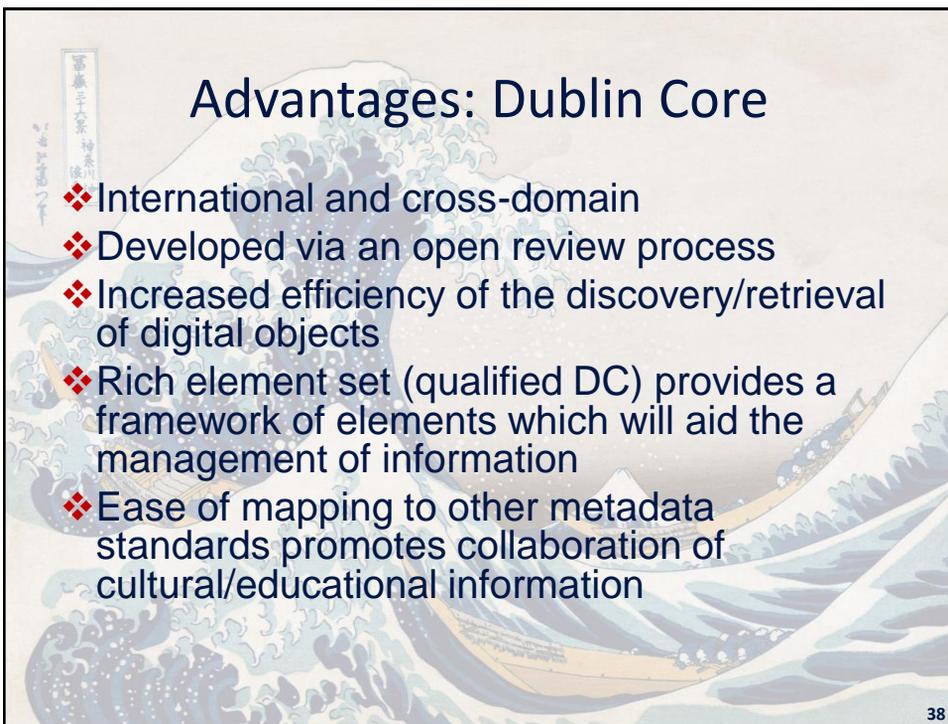
36



## 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 (ES) (Syntax & Vocabulary)
  - Syntax ES: Essentially a datatype that communicates the format or structure of a string
  - Vocabulary ES: Includes values from an identified controlled vocabulary or list

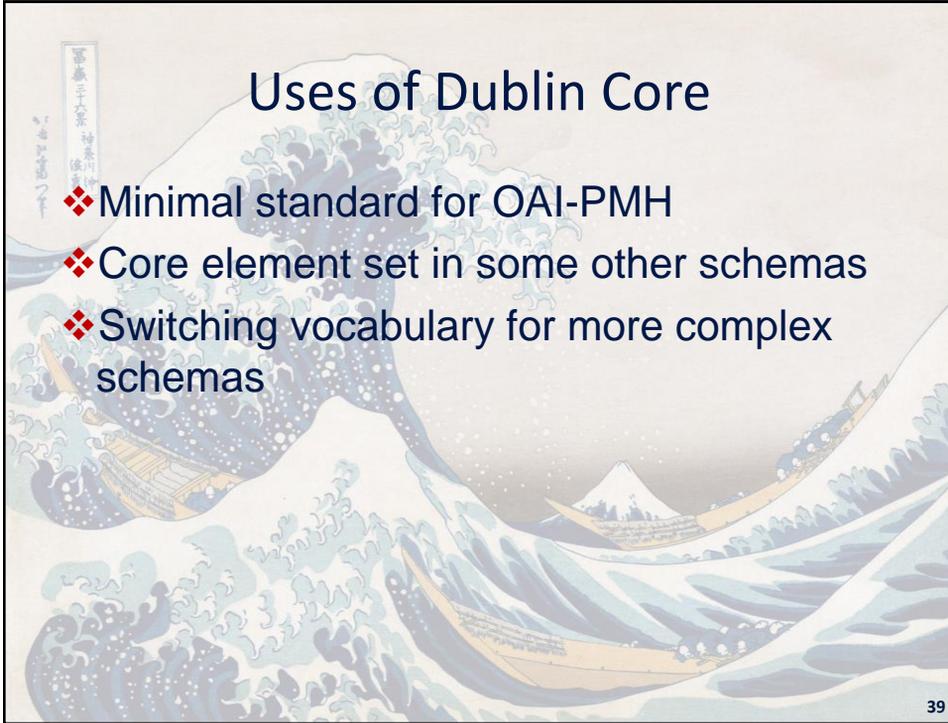
37



## 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

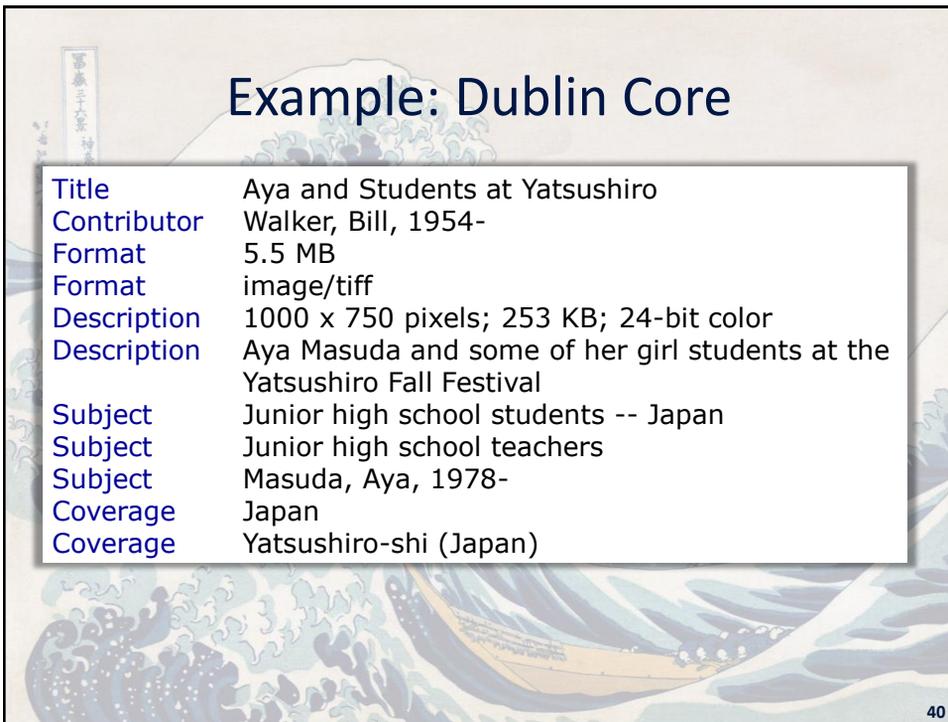
38



## Uses of Dublin Core

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

39



## Example: Dublin Core

Title	Aya and Students at Yatsushiro
Contributor	Walker, Bill, 1954-
Format	5.5 MB
Format	image/tiff
Description	1000 x 750 pixels; 253 KB; 24-bit color
Description	Aya Masuda and some of her girl students at the Yatsushiro Fall Festival
Subject	Junior high school students -- Japan
Subject	Junior high school teachers
Subject	Masuda, Aya, 1978-
Coverage	Japan
Coverage	Yatsushiro-shi (Japan)

40

## Qualified Dublin Core

```

<title qualifier="officialtitle">Fort Worth in the 1920's</title>
<date qualifier="digitized">2007-08-14</date>
<language>nol</language>
<description qualifier="content">Main Street in Fort Worth in the
  1920's</description>
<description qualifier="physical">8x10 b&w glossy
  print</description>
<subject qualifier="KWD">Fort Worth - History</subject>
<subject qualifier="LCSH">Streets - History</subject>
<coverage qualifier="placeName">United States - Texas - Tarrant
  County - Fort Worth</coverage>
<rights qualifier="access">public</rights>
<resourceType>image_photo</resourceType>
<format>image</format>

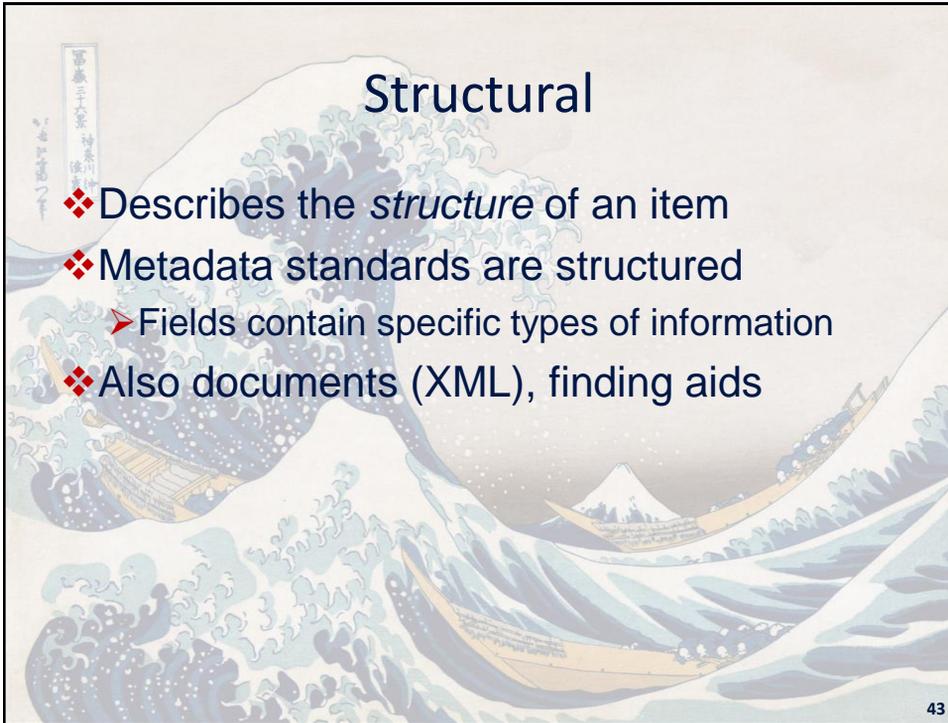
```

41

## Status of DC

- ❖ Dublin Core Metadata Element Set version 1.1
  - ISO Standard 15836-2009; ANSI/NISO Standard Z39.85-2012; 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)

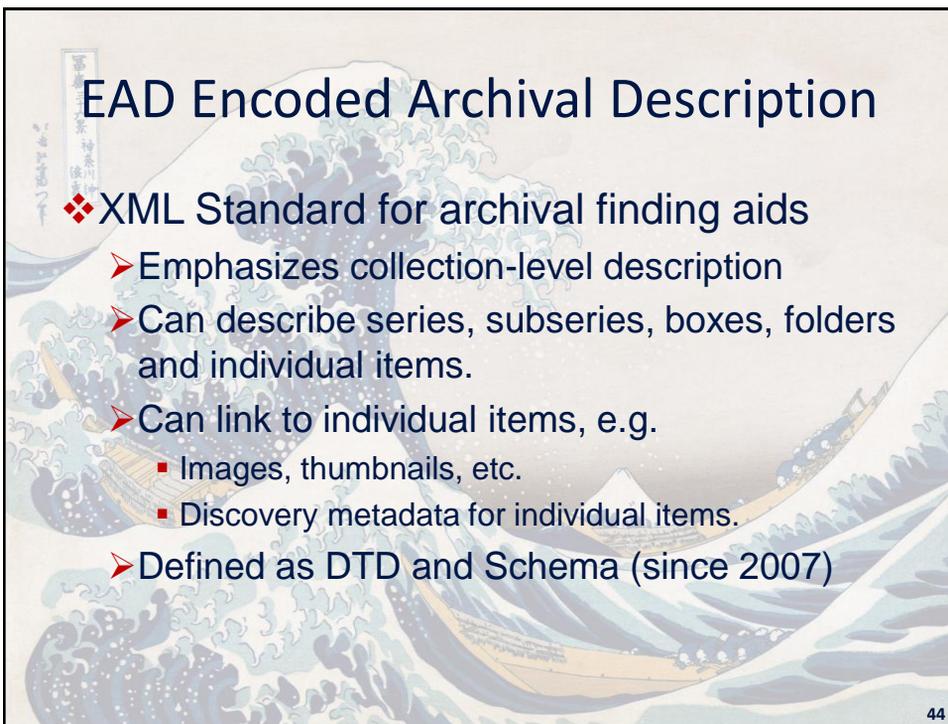
42



## Structural

- ❖ Describes the *structure* of an item
- ❖ Metadata standards are structured
  - Fields contain specific types of information
- ❖ Also documents (XML), finding aids

43



## EAD Encoded Archival Description

- ❖ XML Standard for archival finding aids
  - Emphasizes collection-level description
  - Can describe series, subseries, boxes, folders and individual items.
  - Can link to individual items, e.g.
    - Images, thumbnails, etc.
    - Discovery metadata for individual items.
  - Defined as DTD and Schema (since 2007)

44

# EAD Structure

- ❖ EAD Header
  - Contains information about the finding aid and the institution holding the collection described therein
- ❖ Frontmatter (Optional)
  - A title page for the finding aid
  - (Deprecated in EAD3)
- ❖ Archival description
  - Contains a description of the collection as a whole and, optionally, a multilevel description of the collections parts

45

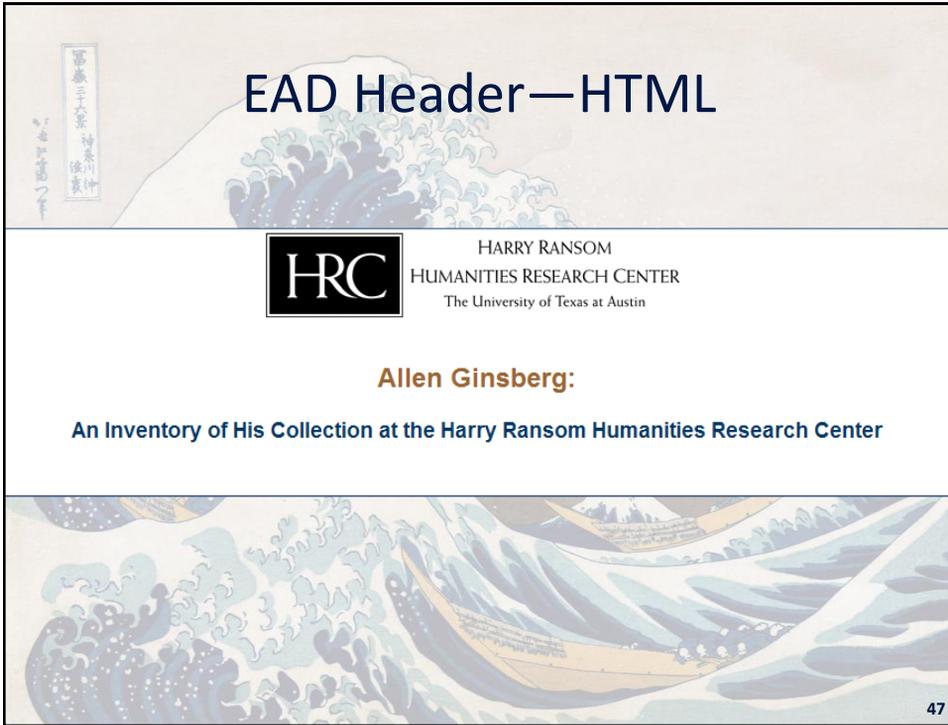
## EAD Header

```

<eadheader langencoding="iso639-2b" audience="internal" id="a0"
repositoryencoding="iso15511" countryencoding="iso3166-1"
scriptencoding="iso15924" dateencoding="iso8601">
  <eadid countrycode="us" mainagencycode="TxU-Hu"
encodinganalog="852$a">urn:taro:utexas.hrc.00044</eadid>
  <filedesc>
    <titlestmt>
      <titleproper>Allen Ginsberg: </titleproper>
      <subtitle>An Inventory of His Collection at the Harry Ransom Humanities
        Research Center</subtitle>
      <author>Chelsea Dinsmore</author>
    </titlestmt>
    <publicationstmt>
      <publisher>University of Texas at Austin</publisher>
      <date era="ce" calendar="gregorian">2000</date>
    </publicationstmt>
  </filedesc>
  <profiledesc>
    <creation>Text converted and initial EAD tagging provided by Apex Data
      Services,
      <date era="ce" calendar="gregorian">September 2000.</date>
    </creation>
    <language>Finding aid written in <language>English.</language>
  </profiledesc>
  <revisiondesc>
    <change>
      <date>Tue Jul 22 15:08:14 CDT 2003</date>
      <item>urn:taro:utexas.hrc.00044 converted from EAD 1.0 to 2002 by
        v1to02.xsl (20030505).</item>
    </change>
  </revisiondesc>
</eadheader>

```

46



# EAD Header—HTML

**HRC** HARRY RANSOM  
HUMANITIES RESEARCH CENTER  
The University of Texas at Austin

**Allen Ginsberg:**  
An Inventory of His Collection at the Harry Ransom Humanities Research Center

47

```

44 <archdesc level="collection">
45 <did id="a1">
46 <head>Descriptive Summary</head>
47 <origination label="Creator">
48 <persname encodinganalog="100">Ginsberg, Allen,
49 1926-1997</persname>
50 </origination>
51 <unittitle label="Title" encodinganalog="245">Allen Ginsberg Collection
52 <unitdate type="inclusive" encodinganalog="245$f" era="ce"
53 calendar="gregorian">1944-1979</unitdate>
54 </unittitle>
55 <unitid countrycode="us" repositorycode="TxU-HU" label="RLIN Record
56 #">TXRC00-A13</unitid>
57 <physdesc label="Extent" encodinganalog="300">3 boxes (1.25 linear feet), 1
58 oversize folder, 2 galley folders, and 12 cassette tapes</physdesc>
59 <repository label="Repository" encodinganalog="852$a">
60 <corpname>
61 <subarea>Harry Ransom Humanities Research Center,
62 </subarea>University of Texas at Austin</corpname>
63 </repository>
64 <abstract>This collection of poet, peace activist, and co-founder of the
65 Beat movement Allen Ginsberg contains a handful of his works, including
66 <title render="italic" linktype="simple">Empty Mirror </title> and a theatrical
67 adaptation
68 of
69 <title render="italic" linktype="simple">Kaddish, </title> correspondence with
70 Peter
71 Orlovsky, Jack Kerouac, and others, and notebooks by and letters to Peter
72 Orlovsky, Ginsberg's long-time companion.</abstract>
73 <langmaterial label="Language">
74 <language langcode="eng">English.</language>
75 </langmaterial>
76 </did>

```

## Descriptive Summary XML View

48

## Descriptive Summary—Web View

### Descriptive Summary

<b>Creator</b>	Ginsberg, Allen, 1926-1997
<b>Title</b>	Allen Ginsberg Collection
<b>Dates:</b>	1944-1979
<b>Abstract:</b>	This collection of poet, peace activist, and co-founder of the Beat movement Allen Ginsberg contains a handful of his works, including <i>Empty Mirror</i> and a theatrical adaptation of <i>Kaddish</i> , correspondence with Peter Orlovsky, Jack Kerouac, and others, and notebooks by and letters to Peter Orlovsky, Ginsberg's long-time companion.
<b>RLIN Record #</b>	TXRC00-A13
<b>Extent</b>	3 boxes (1.25 linear feet), 1 oversize folder, 2 galley folders, and 12 cassette tapes
<b>Language</b>	English.
<b>Repository</b>	Harry Ransom Humanities Research Center, University of Texas at Austin

49

## Scope and contents—XML View

```
<scopecontent id="a3" encodinganalog="520">
  <head>Scope and Contents</head>
  <p>Correspondence and a theatrical adaptation of Ginsberg's poem
  <title render="italic" linktype="simple">Kaddish </title>make up the bulk of the Allen
  Ginsberg Collection, 1944-1979, supplemented by holograph and typescript works
  by Ginsberg, journal and notebook entries by Peter Orlovsky, and critical works
  about Ginsberg by other authors. The collection is organized into four series,
  arranged alphabetically by author and chronologically where possible: Series I.
  Works, 1951-1970 (.5 box); Series II. Correspondence, 1944-1979 (.5 box);
  Series III. Peter Orlovsky's Personal Papers, 1961-1964 (1 box); and Series IV.
  Third-Party Works and Correspondence, 1954-1968 (1 box). This collection was
  previously accessible through a card catalog, but has been recataloged as part
  of a retrospective conversion project.</p>
  <p>The Works Series contains several poems by Ginsberg including
  <title render="doublequote" linktype="simple">The Green Automobile</title> and
  <title render="doublequote" linktype="simple">Primrose Hill Guru,</title> as well as a statement
  to the Senate subcommittee investigating the use of LSD. Additionally, two
  revised typescripts and galley proofs for
  <title render="italic" linktype="simple">Empty Mirror </title>are present. A complete list of
  Ginsberg's works in this collection is provided in the Index of Works at the
  end of this guide.</p>
</scopecontent>
```

50

## Scope and Contents—Web View

### Scope and Contents

Correspondence and a theatrical adaptation of Ginsberg's poem *Kaddish* make up the bulk of the Allen Ginsberg Collection, 1944-1979, supplemented by holograph and typescript works by Ginsberg, journal and notebook entries by Peter Orlovsky, and critical works about Ginsberg by other authors. The collection is organized into four series, arranged alphabetically by author and chronologically where possible: Series I. Works, 1951-1970 (.5 box); Series II. Correspondence, 1944-1979 (.5 box); Series III. Peter Orlovsky's Personal Papers, 1961-1964 (1 box); and Series IV. Third-Party Works and Correspondence, 1954-1968 (1 box). This collection was previously accessible through a card catalog, but has been recataloged as part of a retrospective conversion project.

The Works Series contains several poems by Ginsberg including "The Green Automobile" and "Primrose Hill Guru," as well as a statement to the Senate subcommittee investigating the use of LSD. Additionally, two revised typescripts and galley proofs for *Empty Mirror* are present. A complete list of Ginsberg's works in this collection is provided in the Index of Works at the end of this guide.

The Correspondence Series is divided into two sections, outgoing and incoming. The outgoing section contains a great many letters from Ginsberg to fellow Beat writer Jack Kerouac and Ginsberg's long-time companion Peter Orlovsky, in addition to other friends and writers. The incoming correspondence includes a few letters from consular officials as well as friends and admirers. Individual correspondents are listed in the Index of Correspondents at the end of this guide.

The Peter Orlovsky Series is composed of various bills, calling cards, notes, drafts and other fragments of Orlovsky's professional efforts. Several pages of journal entries are present in addition to an address book. People writing to Orlovsky include publishers, friends, and family. Due to their fragility, the journal pages are only available to researchers in photocopy form. Individual correspondents are listed in the Index of Correspondents.

The Third-party Works and Correspondents Series contains several critical works about Ginsberg's poetry, a few unidentified poems, and multiple drafts of Jerome Benjamin's efforts to adapt Ginsberg's long poem *Kaddish* for the theatre. There is a small amount of correspondence between friends of Ginsberg and Orlovsky, generally about one or the other of the two. Individual authors and their works are listed in the Index of Works by other Authors and individual correspondents are listed in the Index of Correspondents at the end of this guide.

Located elsewhere in the Ransom Center are five Vertical Files containing newspaper clippings with biographical information, literary criticism, and published works by Ginsberg in addition to 32 photographs of Ginsberg, his family, and friends located in the Literary Files of the Photography Collection. Also present are 12 cassette tapes of Ginsberg reading his poetry, and Ginsberg material is located on side B of Gerard Malanga reel-to-reel tape #19.

51

```
<dsc type="in-depth" id="a23">
  <head>Allen Ginsberg Collection--Folder List</head>
  <c01 level="series" id="ser1">
    <did>
      <unittitle>Series I. Works,
    <unitdate era="ce" calendar="gregorian">1951-1970</unitdate>
    </unittitle>
    </did>
    <c02>
      <did>
        <container type="box">1</container>
        <container type="folder">1</container>
        <unittitle>A-Z</unittitle>
      </did>
    </c02>
    <c02>
      <did>
        <container type="box">1</container>
        <container type="folder">2</container>
        <unittitle>
          <title render="italic" linktype="simple">Empty Mirror, </title>
          <unitdate era="ce" calendar="gregorian">1961, </unitdate>
        </unittitle>
        <physdesc>two typescripts with revisions and front matter, 51pp
        (galley proofs removed to galley folder 1)</physdesc>
      </did>
    </c02>
  </c01>
  [ ... ]
</dsc>
```

## Container List XML View

52

## Container List—Web View

### Allen Ginsberg Collection--Folder List

#### Series I. Works, 1951-1970

box	folder
1	1 A-Z
	2 <i>Empty Mirror</i> , 1961, two typescripts with revisions and front matter, 51pp (galley proofs removed to galley folder 1)

[Return to the Table of Contents](#)

53

## TEI Lite

- ❖ Derived from full TEI Guidelines
- ❖ Contains most of TEI “Core” tag set’
- ❖ Adequately handle large variety of texts
  - As demonstrated in the Oxford Text Archive
- ❖ Useful for production of new texts as well as encoding old ones.
- ❖ Small, simple
- ❖ Current version: P5

54

# Header

```

<TEI.2 id="Poem">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>The Sick Rose</title>
        <author>William Blake</author>
      </titleStmt>
      <publicationStmt>
        <pubPlace>Dallas</pubPlace>
        <publisher>Amigosville Community College</publisher>
        <idno>1</idno>
        <availability>
          <p>Available to anyone</p>
        </availability>
        <date value="2014-08-08">August 8, 2014</date>
      </publicationStmt>
    </fileDesc>
  </teiHeader>

```

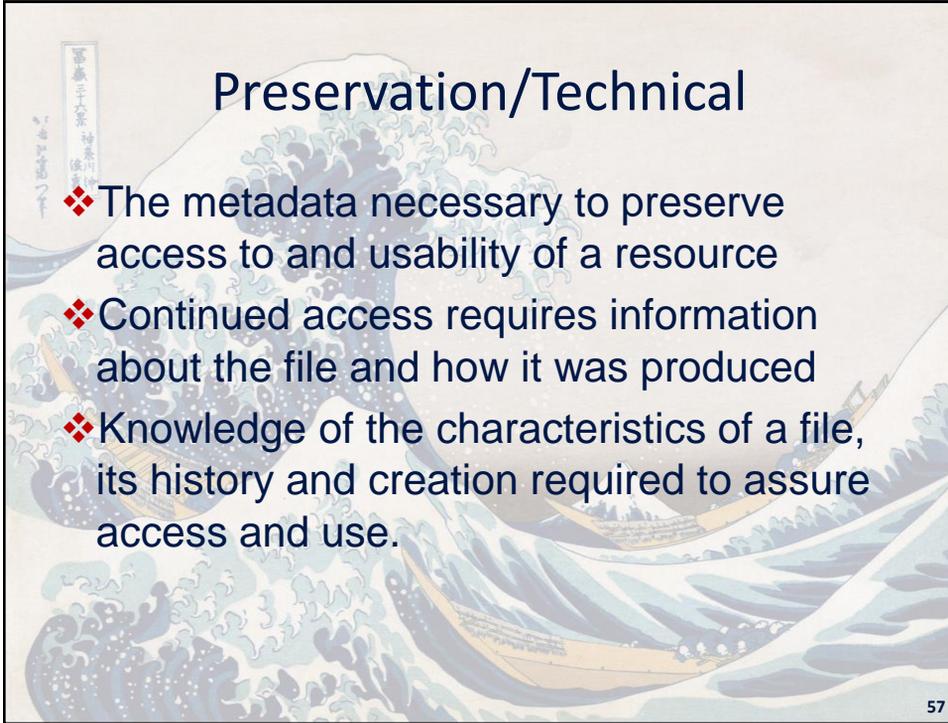
55

```

<text>
  <body>
    <div1>
      <head>
        <title>The SICK ROSE</title>
      </head>
      <lg type="Stanza">
        <l>O Rose thou art sick.</l>
        <l>The invisible worm,</l>
        <l>That flies in the night</l>
        <l>In the howling storm:</l>
      </lg>
      <lg type="Stanza">
        <l>Has found out thy bed</l>
        <l>Of crimson joy:</l>
        <l>And his dark secret love</l>
        <l>Does thy life destroy.</l>
      </lg>
    </div1>
  </body>
</text>
</TEI.2>

```

56



## Preservation/Technical

- ❖ The metadata necessary to preserve access to and usability of a resource
- ❖ Continued access requires information about the file and how it was produced
- ❖ Knowledge of the characteristics of a file, its history and creation required to assure access and use.

57



## Administrative

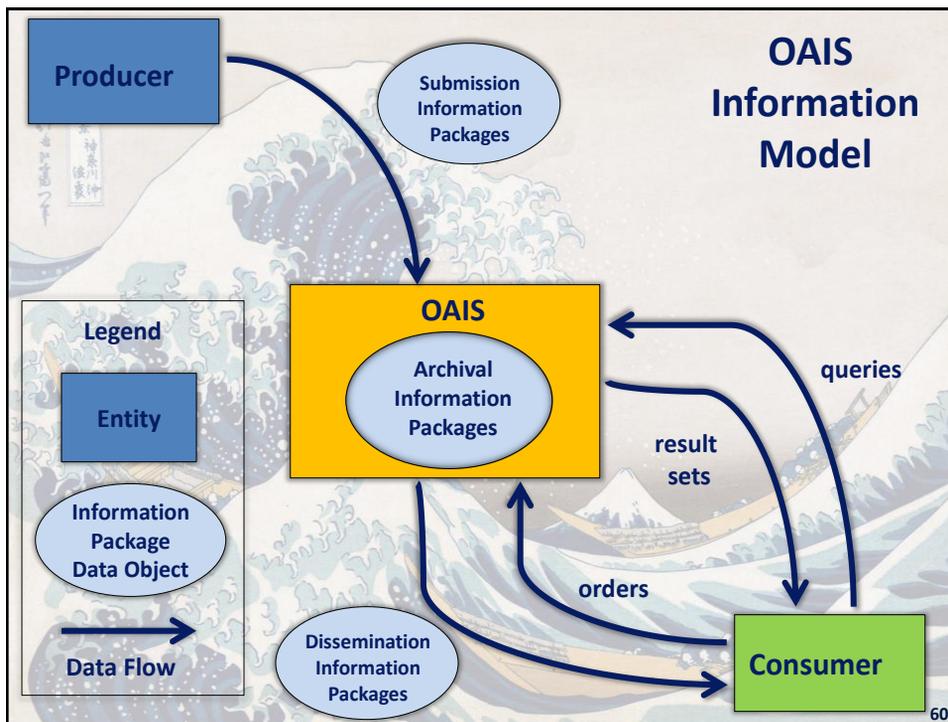
- ❖ Preservation metadata
- ❖ Technical metadata

58

## Open Archival Information System (OAIS) Reference Model

- ❖ Models the components and processes common to most digital preservation activities
- ❖ Developed under the auspices of NASA Consultative Committee for Space Data Systems (CCSDS)
  - ISO standard

59

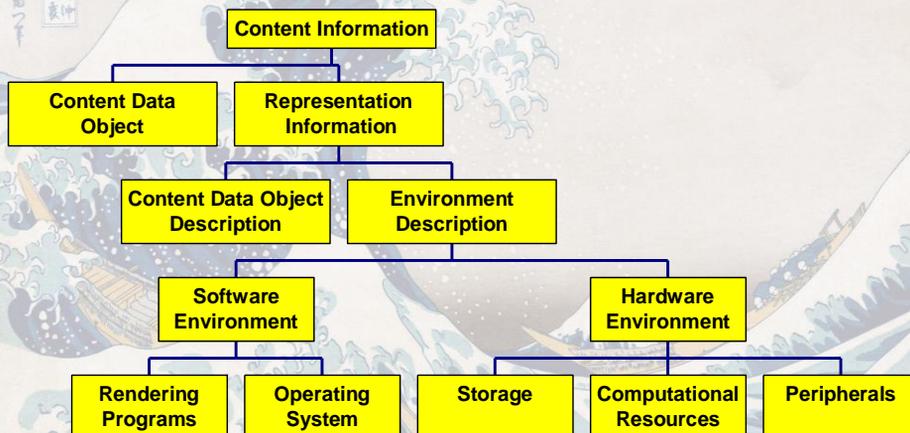


## OAIS Information Model

- ❖ SIP—sent from producer/creator
- ❖ AIP—stored on archive
  - Content information
    - Actual content of object
  - Preservation description information
    - information necessary for preserving object
  - Packaging information
    - binds object with its metadata
  - Descriptive information
    - provides access to object via searching
- ❖ DIP—presented to end user

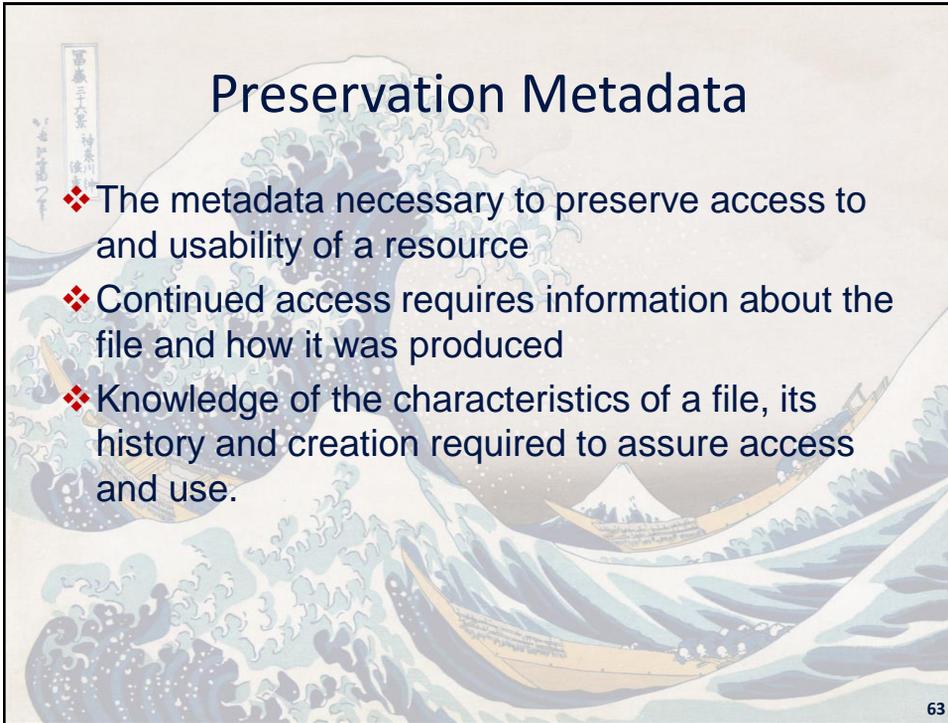
61

## Content Information Components



[http://www.oclc.org/research/projects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/projects/pmwg/pm_framework.pdf)

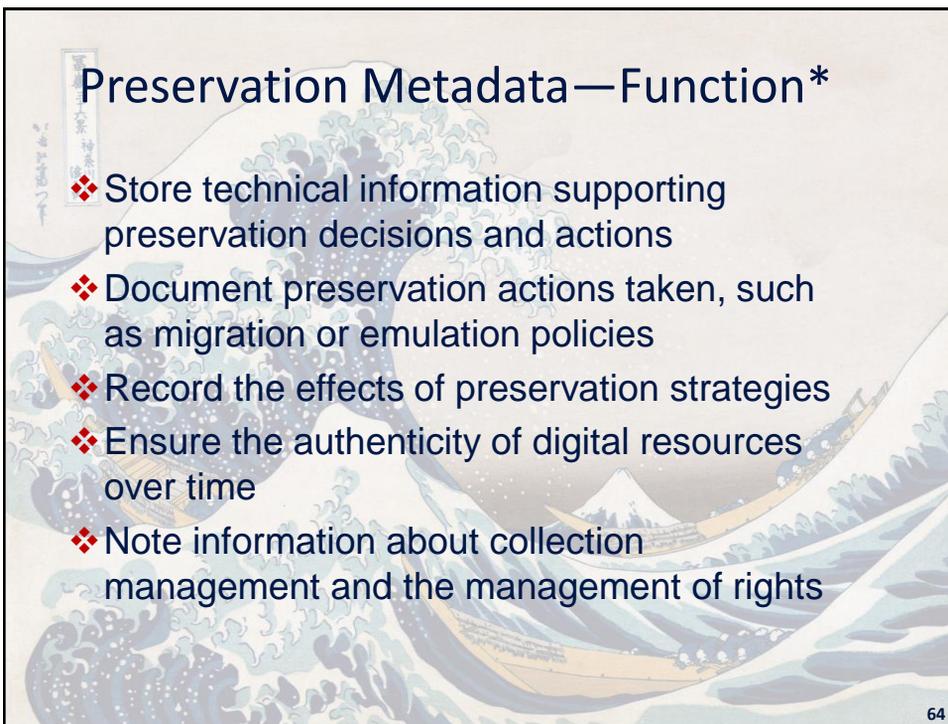
62



## Preservation Metadata

- ❖ The metadata necessary to preserve access to and usability of a resource
- ❖ Continued access requires information about the file and how it was produced
- ❖ Knowledge of the characteristics of a file, its history and creation required to assure access and use.

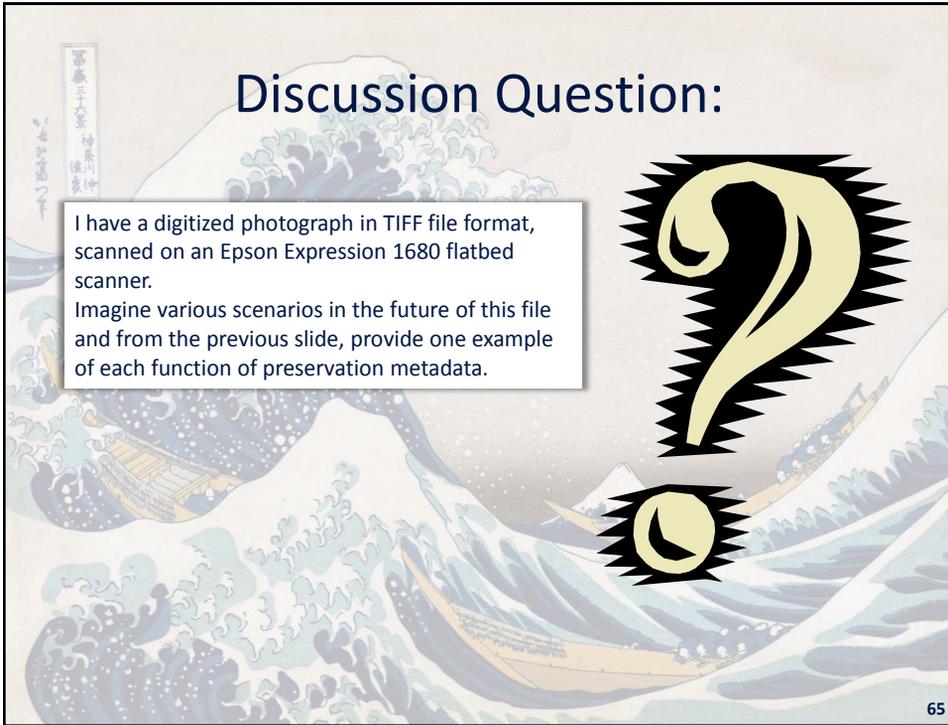
63



## Preservation Metadata—Function\*

- ❖ Store technical information supporting preservation decisions and actions
- ❖ Document preservation actions taken, such as migration or emulation policies
- ❖ Record the effects of preservation strategies
- ❖ Ensure the authenticity of digital resources over time
- ❖ Note information about collection management and the management of rights

64



## Discussion Question:

I have a digitized photograph in TIFF file format, scanned on an Epson Expression 1680 flatbed scanner.

Imagine various scenarios in the future of this file and from the previous slide, provide one example of each function of preservation metadata.

65



## Preservation Metadata—Objectives\*

- ❖ Provide preservation managers with sufficient knowledge to take appropriate actions in order to maintain a digital object's bit stream over the long-term.
- ❖ Ensure that the content of an archived object can be rendered and interpreted, in spite of future changes in access technologies.

\*Quoted in: "Preservation Metadata for Digital Objects: A Review of the State of the Art, op. cit., p. 4.

66

## Preservation Metadata Examples

### ❖ PREMIS

- [www.loc.gov/standards/premis](http://www.loc.gov/standards/premis)

### ❖ NISO

- [http://www.niso.org/standards/resources/Z3987\\_trial\\_use.pdf](http://www.niso.org/standards/resources/Z3987_trial_use.pdf)
- XML schema MIX [www.loc.gov/mix/](http://www.loc.gov/mix/)

### ❖ Audio & Video metadata at LC

67

## PREMIS

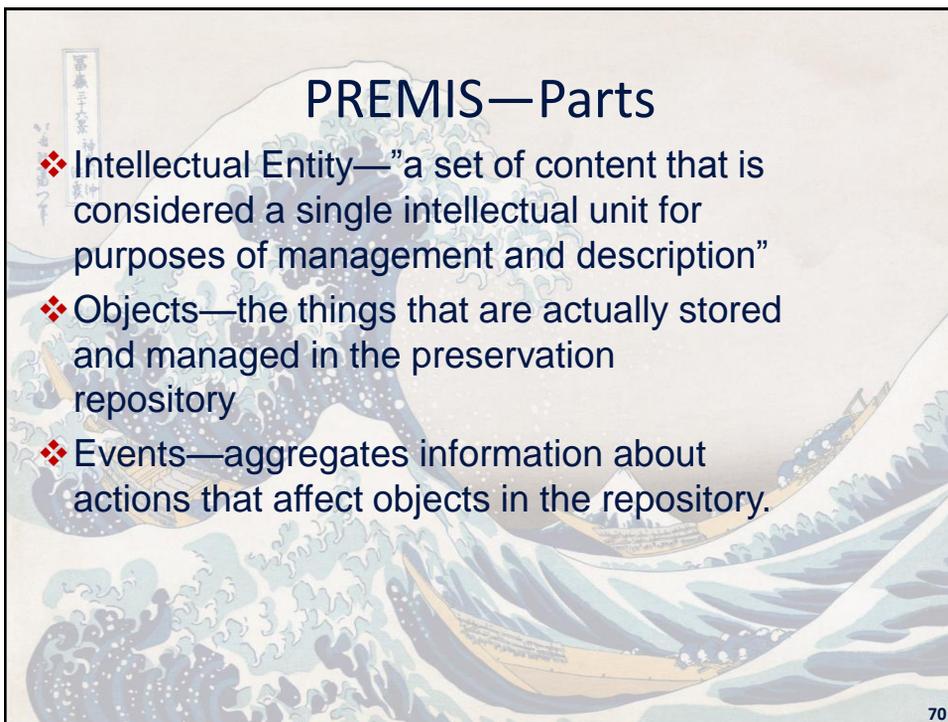
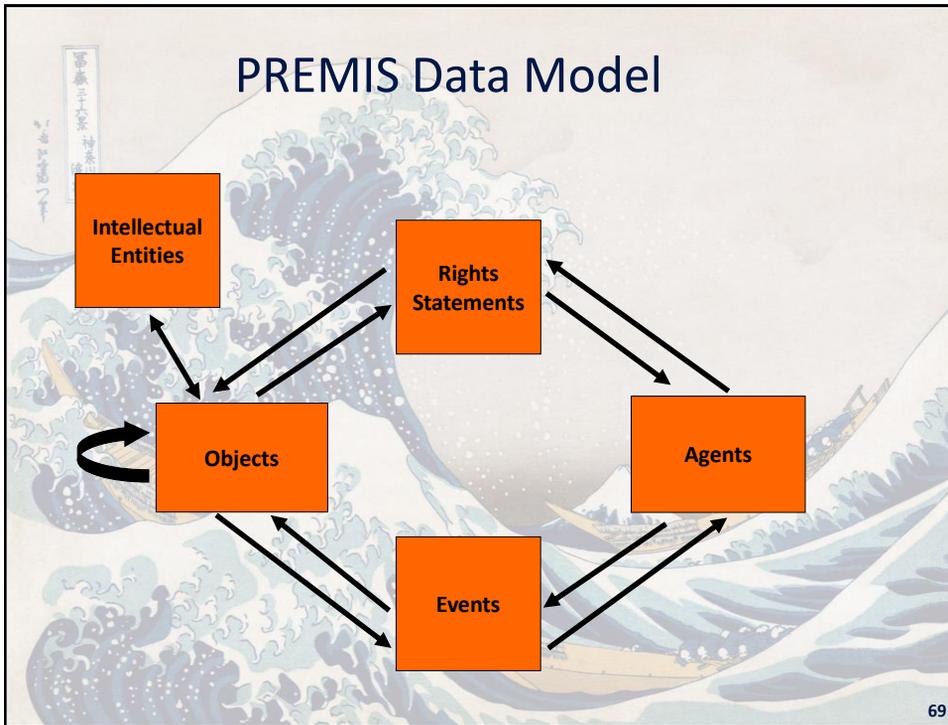
### ❖ PREServation Metadata Implementation Strategies

### ❖ Joint effort between OCLC & RLG

### ❖ Objectives

- Develop a core preservation metadata set, supported by a data dictionary, with broad applicability across the digital preservation community.
- Identify and evaluate alternative strategies for encoding, storing, and managing preservation metadata in digital preservation systems.

68



## PREMIS—Parts

- ❖ Agents—actors that have roles in events and in rights statements
  - ▲ People organizations or software applications
- ❖ Rights—aggregates information about rights and permissions that are directly relevant to preserving objects in the repository.
  - ▲ Acts that the repository has rights to perform
  - ▲ Basis for claiming that right

71

## Audio & Video Technical Metadata

- ❖ Sponsored by LC
  - <http://www.loc.gov/standards/amdvmd/index.html>
- ❖ Audio metadata schemas and documentation by LC (2002) and AES (2011)
- ❖ Video metadata schema by LC

72

## 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

73

## Rights schemas

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

74

## METSRights Metadata Example

```
<?xml version="1.0" encoding="UTF-8"?>
<rts:RightsDeclarationMD xmlns:rts="http://cosimo.stanford.edu/sdr/metsrights/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://cosimo.stanford.edu/sdr/metsrights/
  http://cosimo.stanford.edu/sdr/metsrights.xsd">
  <rts:RightsDeclaration RIGHTSCATEGORY="COPYRIGHTED">
    <rts:Context CONTEXTCLASS="GENERAL PUBLIC">
      <rts:Constraints CONSTRAINTTYPE="RE-USE">
        <rts:ConstraintDescription>The photographs included in this
          collection are copyright Bill Walker, Imaging Field Services
          Officer, Amigos Library Services. Permission to use them for
          the purpose of the "Introduction to METS" workshop is hereby
          granted to registrantos of this workshop. All other uses are
          prohibited without permission of the copyright
          holder.
        </rts:ConstraintDescription>
      </rts:Constraints>
    </rts:Context>
  </rts:RightsDeclaration>
</rts:RightsDeclarationMD>
```

75

## Rights schemas

- ❖ 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 work at the California Digital Library)

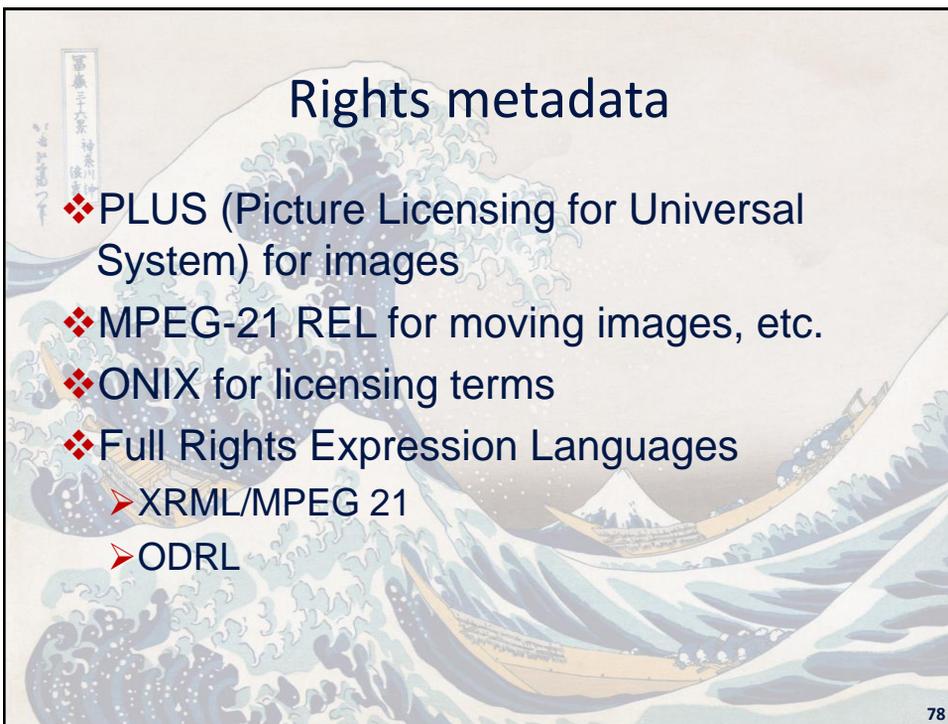
76



## Rights schemas

- ❖ PREMIS Rights
  - Focused on rights for preservation rather than access
  - Revision of PREMIS data dictionary expanded this area
  - Allows for extensibility, i.e. inserting another rights schema
- ❖ Creative Commons
  - Allows creators to choose a license for their work
  - Simple rights statements that fit a lot of situations
  - <http://creativecommons.org/>
  - An example: [MIC catalog](#)

77



## Rights metadata

- ❖ PLUS (Picture Licensing for Universal System) for images
- ❖ MPEG-21 REL for moving images, etc.
- ❖ ONIX for licensing terms
- ❖ Full Rights Expression Languages
  - XRML/MPEG 21
  - ODRL

78



## Assignment

- ❖ Research one metadata standard of your choice
  - One we covered in class
  - One you find elsewhere
- ❖ You can find many standards at: [www.loc.gov/standards](http://www.loc.gov/standards)
- ❖ Provide the following:
  - Brief summary of purpose and objectives of the standard (i.e., why was it created)
  - Tell me the current version of the standard
  - Where can I find documentation for the standard?
  - Copy of the schema (in browser, choose file>>save page as)
- ❖ More details on workshop portal

80