XML Validation
& Schema Languages

Timothy W. Cole & Myung-Ja Han
University of Illinois at Urbana-Champaign

2015 ALA Annual Conference
ALCTS Preconference
25 June 2015
San Francisco, California
Issues When Using XML for Metadata

• Metadata Record Consistency:
  • Optional versus required elements
  • Consistent use of elements
  • Granularity & depth of information

• Interoperability for machine processing:
  • Record to record within a collection or across a CMS
  • Record to record across institutions
As an XML Encoding of the Authors of a Textbook, 
*which of these is best, which worst, why?*

```xml
<authors>
  <auth><last>Cole</last>, <first>Timothy W.</first></auth> and
  <auth><last>Foulonneau</last>, <first>Muriel</first></auth>.
</authors>

```xml
<authors>
  <auth seq='2'><first>Muriel</first><last>Foulonneau</last></auth>
  <auth seq='1'><first>Timothy W.</first><last>Cole</last></auth>
</authors>

```xml
<authors>
  <last>Cole</last> <first>Timothy W.</first>
  <last>Foulonneau</last> <first>Muriel</first>
</authors>
```
Define Document Classes Using DTD or Schema

Formal descriptions of document structure & semantics

• Document Type Definitions (DTDs)
  • Defined in XML spec

• XML Schemas
  • W3C XML Schema Language
  • Schematron
  • Relax NG

• DTDs & Schemas allow us to:
  • Set expectations
  • Maximize reusability
  • Enforce business rules
Valid XML

• Valid XML more work, more rigorous, better for:
  • Larger team projects, with expectations of longevity
  • Heterogeneous, more complex records
  • Expectation that records will be used by multiple different applications (& people) for a variety of purposes
  • Better support for reuse, modular designs
  • But generic XML validation does not replace domain validation!

• XML schema best practices still evolving
  • Attributes versus elements
  • Mixing namespaces
  • Schema languages
  • Philosophical issues
Document Type Definitions (DTD)

• Legacy from SGML; part of XML standard

```xml
<!DOCTYPE Book SYSTEM 'http://...'>
<!ELEMENT Book (Front, Chapter+, Back?)>
<!ATTLIST Book
type (series|monograph) #REQUIRED>
```

• Rarely used in Library contexts today
  (so we’re not going to discuss)
Schemas Part 1 – XML Namespaces

• Qualify element and attribute names
• Namespaces can be tied to a schema
• Allows schema modularization
  • Mix and match elements from multiple schemas in document instances
  • Import or include from one XML Schema into another

```xml
<oai:metadata xmlns:oai='http:...' xmlns:oai_dc='...' xmlns:dc='...'>
  <oai_dc:dc>
    <dc:title>...</dc:title>
    <dc:creator>...</dc:creator>
  </oai_dc:dc>
</oai:metadata>
```
Schemas Part 2 -- schemaLocation

- Schemas are new for XML (not inherited from SGML)
  - Uses XML syntax – schemas are XML document instances
  - Supports datatyping
  - Richer and more complex than DTDs (in some ways)

- W3C XML Schema Language:

```xml
schema for namespace: http://example.edu/bookSchema
  <xsd:element name='Book'>
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name='Front' minOccurs='1' maxOccurs='1' type='frontType'/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>

XML document instance:
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://example.edu/bookSchema/
      http://example.edu/mySchemas/book.xsd"
Alternative languages: Schematron & RelaxNG

• Schematron based on XPath (XSLT)
  • Arguably doesn’t support datatyping as well
  • Supports additional content models
  • May become an ISO standard

• RelaxNG
  • Returns some of the power of SGML DTDs back to XML (mixed and unordered content)
  • Uses datatyping from the XML Schema spec
  • Does not support inheritance
  • Developed by an OASIS Technical Committee chaired by James Clark
  leader="01124cam a2200289 a 4500">
  <controlfield tag="001">3628815</controlfield>
  <controlfield tag="005">20020415165153.0</controlfield>
  <controlfield tag="008">920219s1993 caua j 000 0 eng</controlfield>
  <datafield tag="020" ind1=" " ind2=" ">
    <subfield code="a">0152038655</subfield>
    <subfield code="c">$15.95</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">(OCoLC)ocm25508902</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">AXY-0864</subfield>
  </datafield>
  <datafield tag="040" ind1=" " ind2=" ">
    <subfield code="a">DLC</subfield>
    <subfield code="c">DLC</subfield>
    <subfield code="d">UIU</subfield>
  </datafield>
  <datafield tag="050" ind1="0" ind2="0">
    <subfield code="a">PS3537.A618</subfield>
    <subfield code="b">A88 1993</subfield>
  </datafield>
  <datafield tag="082" ind1="0" ind2="0">
    <subfield code="a">811/.52</subfield>
    <subfield code="c">20</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Sandburg, Carl</subfield>
    <subfield code="d">1878-1967.</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">Arithmetic</subfield>
    <subfield code="c">Carl Sandburg; illustrated as an anamorphic adventure by Ted Rand.</subfield>
  </datafield>
</record>
  <titleInfo>
    <title>Arithmetic</title>
  </titleInfo>
  <name type="personal" usage="primary">
    <namePart>Sandburg, Carl</namePart>
    <namePart type="date">1878-1967</namePart>
  </name>
  <name type="personal">
    <namePart>Rand, Ted</namePart>
  </name>
  <typeOfResource>text</typeOfResource>
  <originInfo>
    <place>
      <placeTerm type="code" authority="iso639-2b" type="text">San Diego</placeTerm>
    </place>
    <publisher>Harcourt Brace Jovanovich</publisher>
    <dateIssued>c1993</dateIssued>
    <dateIssued encoding="marc">1993</dateIssued>
    <edition>1st ed.</edition>
    <issuance>monographic</issuance>
  </originInfo>
  <language>
    <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
  </language>
  <physicalDescription>
    <form authority="marcform">print</form>
    <extent>1 v. (unpaged) : ill. (some col.) ; 26 cm.</extent>
  </physicalDescription>
</mods>
Invalid library metadata records

• An XML-valid MARCXML record *
  http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise2/XML-MARC-MODS-DC/3628815MARC.xml

• An XML-invalid MARCXML record
  http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise1.5/Sample3-MARC-InValid.xml

• MODS XML Schema (version 3.6)
  http://www.loc.gov/standards/mods/v3/mods3-6.xsd

• An XML-invalid MODS XML record
  http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise1.5/Sample3-MODS-InValid.xml

* But compare to results of MARC Validation using MARCEdit...
<?xml version="1.0" encoding="utf-8"?>
  <leader>01124cam a2200289 a 4500</leader>
  <controlfield tag="001">3628815</controlfield>
  <controlfield tag="005">20020415165153.0</controlfield>
  <controlfield tag="008">920219s1993 caua j 000 0 eng</controlfield>
  <datafield tag="020" ind1=" " ind2=" ">
    <subfield code="a">0152038655 :</subfield>
    <subfield code="c">$15.95</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">(OCoLC)ocm25508902</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="9">AXY-0864</subfield>
  </datafield>
  <datafield tag="040" ind1=" " ind2=" ">
    <subfield code="a">DLC</subfield>
    <subfield code="c">DLC</subfield>
    <subfield code="d">UIU</subfield>
  </datafield>
  <datafield tag="050" ind1="0" ind2="0">
    <subfield code="a">PS3537.A618</subfield>
    <subfield code="b">A88 1993</subfield>
  </datafield>
  <datafield tag="082" ind1="0" ind2="0">
    <subfield code="a">811/.52</subfield>
    <subfield code="2">20</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Sandburg, Carl</subfield>
    <subfield code="d">1878-1967</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">Arithmetic /</subfield>
    <subfield code="e">Carl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.</subfield>
  </datafield>
  <datafield tag="250" ind1=" " ind2=" ">
    <subfield code="a">1st ed.</subfield>
  </datafield>
  <datafield tag="250" ind1=" " ind2=" ">
    <subfield code="a">San Diego :</subfield>
    <subfield code="b">Harcourt Brace Jovanovich</subfield>
  </datafield>
</record>
A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.
Illustration of an INVALID MARC21XML Record

Available online

http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise1.5/Sample3-MARC-InValid.xml

<?xml version="1.0" encoding="utf-8"?>
<record xmlns="http://www.loc.gov/MARC21/slim"
xsi:schemaLocation="http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"
xmns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<leader>04070cam a2200505 i 4500</leader>
<controlfield tag="001">7285233</controlfield>
<controlfield tag="005">20130829200348.0</controlfield>
<controlfield tag="008">130327s2013 caua b 001 0 eng</controlfield>
<datafield tag="010" ind1="" ind2="">
$subfield code="a">2013006978</subfield>
</datafield>
<datafield tag="035" ind1="" ind2="">
$subfield code="a">(OCoLC)ocn816512986</subfield>
</datafield>
<datafield tag="040" ind1="" ind2="">
$subfield code="a">DLC</subfield>
$subfield code="b">eng</subfield>
$subfield code="e">rda</subfield>
$subfield code="e">DLC</subfield>
$subfield code="d">YDX</subfield>
$subfield code="d">YDXCP</subfield>
$subfield code="d">BTCTA</subfield>
$subfield code="d">BDX</subfield>
$subfield code="d">OCLCO</subfield>
$subfield code="d">CHVBK</subfield>
$subfield code="d">MUU</subfield>
</datafield>
<datafield tag="020" ind1="" ind2=""/>
<datafield tag="019" ind1="" ind2="">
$subfield code="a">817537028</subfield>
</datafield>
<datafield tag="020" ind1="" ind2="">
$subfield code="a">9781598845198 (pbk. : acid-free paper)</subfield>
</datafield>
<datafield tag="020" ind1="" ind2="">
$subfield code="a">1598845195 (pbk. : acid-free paper)</subfield>
</datafield>
<datafield tag="020" ind1="" ind2="">
$subfield code="z">9781610692915 (ebook)</subfield>
</datafield>
<datafield tag="042" ind1="" ind2="">
$subfield code="a">pcc</subfield>
</datafield>
<datafield tag="050" ind1="0" ind2="0">
$subfield code="a">MARC21slim</subfield>
</datafield>
</record>
This book provides a foundation of knowledge for catalogers, metadata librarians, and library school students on the Extensible Markup Language (XML)--one of the most commonly listed qualifications in today's cataloger and metadata librarian job postings. How are today's librarians to manage and describe the ever-expanding volumes of resources, in both digital and print formats? The use of XML in cataloging and metadata workflows can improve metadata quality, the consistency of cataloging workflows, and adherence to standards. This book is intended to enable current and future catalogers and metadata librarians to progress beyond a bare surface-level acquaintance with XML, thereby enabling them to integrate XML technologies more fully into their cataloging workflows. Building on the wealth of work on library descriptive practices, cataloging, and metadata, XML for Catalogers and Metadata Librarians explores the use of XML to serialize, process, share, and manage library catalog and metadata records. The authors' expert treatment of the topic is written to be accessible to those with little or no prior practical knowledge of or experience with how XML is used. Readers will gain an educated appreciation of the nuances of XML and grasp the benefit of more advanced and complex XML techniques as applied to applications relevant to catalogers and metadata librarians."--Publisher's website.

<datafield tag="650" ind1="" ind2="" >
  <subfield code="a">XML.</subfield>
  <subfield code="2">gnd</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="7" >
  <subfield code="a">Katalogisierung.</subfield>
  <subfield code="2">gnd</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="7" >
  <subfield code="a">Metadaten.</subfield>
  <subfield code="2">gnd</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="7" >
  <subfield code="a">Datenverarbeitung.</subfield>
  <subfield code="2">gnd</subfield>
</datafield>
<datafield tag="700" ind1="1" ind2="" >
  <subfield code="a">Han, Myung-Ja K.</subfield>
</datafield>
</record>
Illustration of an **INVALID MODS XML Record**

Available online

http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise1.5/Sample3-MODS-Invalid.xml

```xml
<?xml version="1.0" encoding="utf-8"?>
<mods version="3.4"
  <titleInfo>
    <title>Using the Open Archives Initiative protocol for metadata harvesting</title>
  </titleInfo>
  <name type="personal" usage="primary">
    <namePart type="family">Cole</namePart>
    <namePart type="given">Timothy W.</namePart>
  </name>
  <typeOfResource>text</typeOfResource>
  <name>Foulonneau, Muriel.</name>
  <genre authority="marcgt">bibliography</genre>
  <originInfo>
    <placeTerm type="code" authority="marccountry">ctu</placeTerm>
    <place>
      <placeTerm type="text">Westport, Conn</placeTerm>
    </place>
    <publisher>Libraries Unlimited</publisher>
    <dateIssued>c2007</dateIssued>
    <dateIssued encoding="marc">2007</dateIssued>
    <issuance>monographic</issuance>
  </originInfo>
  <language>
    <languageTerm authority="iso639-2b" type="Code">eng</languageTerm>
  </language>
  <physicalDescription>
    <form authority="marcform">print</form>
    <extent>xv, 208 p. : ill. ; 26 cm.</extent>
  </physicalDescription>
  <tableOfContents>
    Definition and origins of OAI-PMH -- Underlying technologies and the technical development of OAI-PMH -- Context for OAI-PMH: eprints, institutional repositories, and open access -- Technical details of the protocol -- Implementing an OAI data provider -- Creating metadata to share -- Post-harvest metadata normalization & augmentation -- Using aggregated metadata to build digital library services -- Concluding thoughts.
  </tableOfContents>
  <note type="bibliography">Includes bibliographical references and index.</note>
  <subject authority="lcsh">
    <name type="corporate">
      <namePart>Open Archives Initiative.</namePart>
    </name>
  </subject>
  <subject authority="lcsh">
    <topic>Metadata harvesting</topic>
  </subject>
</mods>
```
Metadata Crosswalk
So Many Metadata Standards...

Metadata Conversion

Transform one metadata standard to another standard or extract information from metadata
Three Steps

• **Semantic Interoperability**
  Enables you to provide consistent services across multiple collections of resources even when the semantics used to describe items in these collections vary by collection

• **Metadata Crosswalks**
  Maps that describe semantic equivalences of elements in one schema to elements in another; typically expressed as tables

• **Metadata Transforms**
  Implementations of metadata crosswalks; typically accomplished using XML Transforming Stylesheets (XSL-T)
How to Convert Metadata?

• **Mapping**
  • The intellectual activity of comparing and analyzing two or more metadata schemas

• **Crosswalk**
  • The visual and textual product of the mapping process
# MARC to DC Crosswalk

<table>
<thead>
<tr>
<th>MARC Data fields</th>
<th>Dublin Core Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>130, 240, 245, 246</td>
<td>Title</td>
</tr>
<tr>
<td>100, 111, 100</td>
<td>Creator</td>
</tr>
<tr>
<td>700, 710, 711</td>
<td>Contributor</td>
</tr>
<tr>
<td>600, 610, 611, 650, 651, 630, 655</td>
<td>Subject</td>
</tr>
<tr>
<td>500, 504, 505, 520, 562, 583...</td>
<td>Description</td>
</tr>
<tr>
<td>Control field 008 location 07-10, 260 $c</td>
<td>Date</td>
</tr>
<tr>
<td>260 $b</td>
<td>Publisher</td>
</tr>
</tbody>
</table>
Metadata in XML

Examples of same metadata in different formats:

- http://quest.library.illinois.edu/GetMARC/one.aspx/3628815.xml
- http://quest.library.illinois.edu/GetMARC/one.aspx/3628815.mods
- http://quest.library.illinois.edu/GetMARC/one.aspx/3628815.dc
Illustration of Metadata Crosswalking
Cached copies available online

http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise2/XML-MARC-MODS-DC/3628815MARC.xml
http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise2/XML-MARC-MODS-DC/3628815MODS.xml

MARC:

<?xml version="1.0" encoding="utf-8"?>
  <leader>01124cam a2200289 a 4500</leader>
  <controlfield tag="001">3628815</controlfield>
  <controlfield tag="005">20020415165153.0</controlfield>
  <controlfield tag="008">920219s1993 caua j 000 0 eng</controlfield>
  <datafield tag="020" ind1=" " ind2=" ">
    <subfield code="a">0152038655 ::</subfield>
    <subfield code="c">$15.95</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">(OCoLC)ocm25508902</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="9">AXY-0864</subfield>
  </datafield>
  <datafield tag="040" ind1=" " ind2=" ">
    <subfield code="a">DLC</subfield>
    <subfield code="c">DLC</subfield>
    <subfield code="d">UIU</subfield>
  </datafield>
  <datafield tag="050" ind1="0" ind2="0">
    <subfield code="a">PS3537.A618</subfield>
    <subfield code="b">A88 1993</subfield>
  </datafield>
  <datafield tag="082" ind1="0" ind2="0">
    <subfield code="a">811/.52</subfield>
    <subfield code="2">20</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2="0">
    <subfield code="a">Sandburg, Carl,</subfield>
    <subfield code="d">1878-1967.</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">Arithmetic</subfield>
  </datafield>
  <datafield tag="250" ind1=" " ind2=" ">
    <subfield code="a">1st ed.</subfield>
  </datafield>
  <datafield tag="260" ind1=" " ind2=" ">
    <subfield code="a">San Diego ;</subfield>
  </datafield>
  <datafield tag="264" ind1=" " ind2=" ">
    <subfield code="b">Harcourt Brace Jovanovich, </subfield>
    <subfield code="c">c1993.</subfield>
  </datafield>
  <datafield tag="300" ind1=" " ind2=" ">
    <subfield code="a">1 v. (unpaged).</subfield>
    <subfield code="b">ill. (some col.).</subfield>
  </datafield>
</record>
A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.

Children's poetry, American.

Arithmetic

American poetry.

Visual perception.

Arithmetic

Juvenile poetry.

Rand, Ted
A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.

Target Audience: Juvenile

Subject: Children's poetry, American, American poetry, Visual perception, Arithmetic, Juvenile poetry

Classification: PS3537.A618 A88 1993, 811/.52

Identifier: ISBN 0152038655, OCLC 25508902

Record Origin: Converted from MARCXML to MODS version 3.4 using MARC21slim2MODS3-4.xsl (Revision 1.77 2012/03/23)

DUBLIN CORE:

<?xml version="1.0" encoding="utf-8"?>
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:oai_dc="http://purl.org/dc/elements/1.1/"
<dc:title>Arithmetic</dc:title>
A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.

One Mylar sheet included in pocket.

A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.

Children's poetry, American.

Arithmetic

American poetry.

Visual perception.

Arithmetic

URN:ISBN:0152038655
Non-alignment between Elements

• **Easy**: elements A & B of schema 1 must be combined to populate element C of schema 2 (many-to-one)

• **Easy**: distinct elements A & B of schema 1 each map to same element C of schema 2

• **Hard**: value of element C of schema 1 must be parsed to populate elements A & B of schema 2 (one-to-many)

• **Hard**: ambiguity -- value in element C of schema 1 could go into either element A or element B of schema 2
### For Example

<table>
<thead>
<tr>
<th>Schema 1</th>
<th>Schema 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODS</td>
<td>Dublin Core</td>
</tr>
<tr>
<td>&lt;name type=&quot;personal&quot;&gt;</td>
<td>&lt;creator&gt;</td>
</tr>
<tr>
<td>&lt;namePart type=&quot;given&quot;&gt;Jack&lt;/namePart&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;namePart type=&quot;family&quot;&gt;May&lt;/namePart&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;namePart type=&quot;date&quot;&gt;1947-&lt;/namePart&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;name&gt;</td>
<td></td>
</tr>
<tr>
<td>MARC datafields 100, 110, 111</td>
<td>Dublin Core</td>
</tr>
<tr>
<td>&lt;creator&gt;</td>
<td>&lt;creator&gt;</td>
</tr>
<tr>
<td>Dublin Core</td>
<td>MODS? MARC?</td>
</tr>
<tr>
<td>&lt;creator&gt;</td>
<td></td>
</tr>
<tr>
<td>MARC datafields 700, 710, 711, 720</td>
<td>Dublin Core</td>
</tr>
<tr>
<td>&lt;creator&gt; or &lt;contributor&gt;?</td>
<td></td>
</tr>
</tbody>
</table>
Granularity and Completeness

- Some elements in schema 1 may be less fine-grained than analogous elements in schema 2
- Elements in schema 1 may have no compatible elements in schema 2
- Element required by schema or application profile for schema 2 may be optional or missing from schema 1
  - DC coverage simultaneously encompasses both MODS subject-temporal and MODS subject-geographic
  - MODS <targetAudience> has no analog in Simple DC
  - DLF Aquifer MODS guidelines require <recordInfo> for which there is no analog in Simple DC
World View of Schemas may Differ

- DC 1-to-1 Principle vs. VRA Core
- EAD hierarchical description vs. DC flat item-level description
- MARC bibliographic description vs. Darwin Core Taxonomic approach (http://rs.tdwg.org/dwc/terms/)
Crosswalk Implementations

• A crosswalk is lossless if you can crosswalk a schema 1 record into schema 2 record and then back into schema 1 record

• **However**, in the real world, **most crosswalks are lossy**

• Typically, less information is lost when going from a rich metadata scheme into a less rich metadata scheme

• To implement a crosswalk, need to turn mappings into **rules** and then implement these rules, e.g., through a transforming XML Stylesheet (XSL-T)
Exercise 2 -- Create valid MODS records in XML
http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise2/

Help with MODS – element & attribute definitions & guidance:
http://www.loc.gov/standards/mods/userguide/generalapp.html &

• Start with the well-formed metadata you created in Exercise 1
  More detailed instructions

• Template4Exercise2
  http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise2/Template4Exercise2.xml
Exercise 2
Creating valid MODS metadata in XML

1. Please create valid MODS metadata for as many of the well-formed XML documents that you created in Exercise 1 as you can. Use the following MODS root element which includes reference to the current MODS XML Schema. See also starting template for Exercise 2:
   http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise2/Template4Exercise2.xml

   * Definitions of all MODS Top elements, subelements, and attributes are available here:
     http://www.loc.gov/standards/mods/userguide/generalapp.html
   See also this guide:

   * MODS metadata header:

     <mods
     xmlns:xlink="http://www.w3.org/1999/xlink"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xmlns="http://www.loc.gov/mods/v3"

     For your first attempt, use only the following top-level MODS elements with as many children elements of these as you want.

     <titleInfo>
     <name>
     <typeOfResource>
     <subject>
     <originInfo>

     Be sure to validate your MODS records in oXygen.

2. If you have time, identify additional MODS Top elements, sub-elements, and attributes that can be used for creating complete MODS metadata for all items.

3. You can also compile all MODS metadata as one MODS Collection metadata. Please think about how to do this.
Valid XML Exercise
A Starting Template for Exercise 2

Available online
http://quest.library.illinois.edu/ALA2015/ALCTS-XML/Exercises/Exercise2/Template4Exercise2.xml

<?xml version="1.0" encoding="utf-8"?>
<mods version="3.6"
  <titleInfo>
    <title></title>
  </titleInfo>

  <name type="personal" usage="primary">
    <namePart type="family">Cole</namePart>
    <namePart type="given">Timothy W.</namePart>
    <role>
      <roleTerm type="code" authority="marcrelator">aut</roleTerm>
    </role>
  </name>

  <originInfo>
    <place>
      <placeTerm></placeTerm>
    </place>
    <dateCreated></dateCreated>
    <publisher></publisher>
  </originInfo>
</mods>
Exercise 2 Solutions

• Our Solutions:
  Sample1-MODS.xml
  Sample2-MODS.xml
  Sample3-MODS.xml
  Sample4-MODS.xml

• Invalid:
  Sample2-MODS-InValid.xml
  Sample4-MODS-InValid.xml

Discussion: Does having an XML Schema help you make more consistent metadata records?
Valid XML Exercise
Solution for Resource #1

Available online
http://quest.library.illinois.edu/ALA2015/ALCTS-ExerciseSolutions/Exercise2/Valid/Sample1-MODS.xml

<?xml version="1.0" encoding="utf-8"?>
  <titleInfo>
    <title>William T. Young Library, University of Kentucky</title>
  </titleInfo>
  <name type="corporate">
    <namePart>University of Illinois at Urbana-Champaign. University Archives</namePart>
    <role>
      <roleTerm type="text" authority="marcrelator">pul</roleTerm>
      <roleTerm type="code" authority="marcrelator">publisher</roleTerm>
    </role>
  </name>
  <typeOfResource>still image</typeOfResource>
  <accessCondition>Unknown</accessCondition>
  <originInfo>
    <place>
      <placeTerm type="code" authority="marccountry">ilu</placeTerm>
      <placeTerm type="text" authority="marccountry">Illinois</placeTerm>
    </place>
    <dateCreated>October 21 2008</dateCreated>
    <publisher>ALA Archives - University of Illinois at Urbana-Champaign</publisher>
  </originInfo>
  <physicalDescription>
    <form>TIFF</form>
  </physicalDescription>
  <language>
    <languageTerm>English</languageTerm>
  </language>
  <note>On Front: William T. Young Library, University of Kentucky; On Back: The William T. Young Library opened on the University of Kentucky campus in April 1998. It is the third central library in the University's history, replacing the Margaret T. King Library which opened in 1931 and the Carnegie Library constructed in 1909. Michael McKinell of Boston was the architect and Bovis the construction management firm. The building has over 350,000 square feet of space and cost $58 million. President Wethington's commitment to the project was critical as the facility was constructed with private funds by the UK Alumni Association and bonds were sold to finance part of the construction expenses. William T. Young of Lexington was the leading contributor to the Library.</note>
  <subject>
    <topic>Library buildings Kentucky</topic>
  </subject>
</mods>
<subject>
  <topic>Academic libraries</topic>
</subject>

<subject>
  <geographic>Lexington, Kentucky</geographic>
</subject>

<relatedItem type="host" displayLabel="Collection">
  <titleInfo>
    <title>Sjoerd Koopman Library Postcard Collection</title>
  </titleInfo>
  <location displayLabel="Collection Available at">
    <url>http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/koopman</url>
  </location>
</relatedItem>

<relatedItem type="original">
  <physicalDescription>
    <form>paper</form>
    <note>black and white postcard</note>
    <extent>14.0x10.8cm</extent>
  </physicalDescription>
  <location>
    <physicalLocation>University of Illinois at Urbana-Champaign. University Archives</physicalLocation>
  </location>
</relatedItem>

<identifier type="local">ALA0000483</identifier>
<identifier type="uri">http://imagesearchnew.library.illinois.edu/cdm/ref/collection/koopman/id/777</identifier>
Valid XML Exercise
Solution for Resource #2

Available online
http://quest.library.illinois.edu/ALA2015/ALCTS-ExerciseSolutions/Exercise2/Valid/Sample2-MODS.xml

<?xml version="1.0" encoding="utf-8"?>
  <titleInfo>
    <title>50-50: Fighting Chicago's Crime Trusts</title>
  </titleInfo>

  <name type="personal">
    <namePart type="family">French</namePart>
    <namePart type="given">George W.</namePart>
    <role>
      <roleTerm type="text" authority="marcrelator">cre</roleTerm>
      <roleTerm type="code" authority="marcrelator">creator</roleTerm>
    </role>
  </name>

  <typeOfResource>still image</typeOfResource>

  <accessCondition>Images in this collection were digitized through the University of Illinois Library's participation in the Open Content Alliance and may be used freely. Attribution to the University of Illinois is appreciated. High-resolution images can be downloaded from the Internet Archive at www.archive.org. For further information, contact dcc@library.uiuc.edu.</accessCondition>

  <originInfo>
    <place>
      <placeTerm type="code" authority="marccountry">ilu</placeTerm>
      <placeTerm type="text" authority="marccountry">Illinois</placeTerm>
    </place>
    <dateCreated>1916</dateCreated>
  </originInfo>

  <physicalDescription>
    <internetMediaType>image/jp2</internetMediaType>
  </physicalDescription>

  <language>
    <languageTerm>English</languageTerm>
  </language>

  <note>Cover of the book "50-50": Fighting Chicago's Crime Trusts.</note>

  <subject>
    <topic>Book covers</topic>
  </subject>

  <subject>
    <topic>Covers (Illustration)</topic>
  </subject>

  <subject>
    <geographic>Chicago (Ill.)</geographic>
  </subject>
</mods>
Valid XML Exercise
Solution for Resource #3

Available online
http://quest.library.illinois.edu/ALA2015/ALCTS-ExerciseSolutions/Exercise2/Valid/Sample3-MODS.xml

<?xml version="1.0" encoding="utf-8"?>

<titleInfo>
<title>XML for catalogers and metadata librarians</title>
</titleInfo>

<name type="personal" usage="primary">
<namePart type="family">Cole</namePart>
<namePart type="given">Timothy W.</namePart>
<role>
<roleTerm type="text" authority="marcrelator">author</roleTerm>
<roleTerm type="code" authority="marcrelator">aut</roleTerm>
</role>
</name>

<name type="personal">
<namePart type="family">Han</namePart>
<namePart type="given">Myung-Ja K.</namePart>
<role>
<roleTerm type="text" authority="marcrelator">author</roleTerm>
<roleTerm type="code" authority="marcrelator">aut</roleTerm>
</role>
</name>

<typeOfResource>text</typeOfResource>

<originInfo>
<place>
<placeTerm type="code" authority="marccountry">cau</placeTerm>
<placeTerm type="text" authority="marccountry">Santa Barbara, California</placeTerm>
</place>
<dateIssued encoding="marc">2013</dateIssued>
<issuance>monographic</issuance>
<publisher>Libraries Unlimited, an imprint of ABC-CLIO, LLC</publisher>
</originInfo>

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

<physicalDescription>
<form authority="marcform">print</form>
<extent>xiv, 388 pages : illustrations ; 26 cm.</extent>
</physicalDescription>

<abstract>This book provides a foundation of knowledge for catalogers, metadata librarians, and library school students on the Extensible Markup Language (XML)--one of the most commonly listed qualifications in today's cataloger and metadata librarian job postings. How are today's librarians to manage and describe the ever-expanding volumes of resources, in both digital and print formats? The use of XML in cataloging and metadata workflows can improve metadata quality, the consistency of cataloging workflows, and adherence to standards. This book is intended to enable current and future catalogers and metadata librarians to
progress beyond a bare surface-level acquaintance with XML, thereby enabling them to integrate XML technologies more fully into their cataloging workflows. Building on the wealth of work on library descriptive practices, cataloging, and metadata, XML for Catalogers and Metadata Librarians explores the use of XML to serialize, process, share, and manage library catalog and metadata records. The authors’ expert treatment of the topic is written to be accessible to those with little or no prior practical knowledge of or experience with how XML is used. Readers will gain an educated appreciation of the nuances of XML and grasp the benefit of more advanced and complex XML techniques as applied to applications relevant to catalogers and metadata librarians."--Publisher's website.


<subject authority="lcsh">
  <topic>XML (Document markup language)</topic>
</subject>
<subject authority="lcsh">
  <topic>Cataloging</topic>
</subject>
<subject authority="lcsh">
  <topic>Data processing</topic>
</subject>
<subject authority="lcsh">
  <topic>Metadata</topic>
</subject>
<subject authority="bisacsh">
  <topic>LANGUAGE ARTS &amp; DISCIPLINES / Library &amp; Information Science / General</topic>
</subject>
<subject authority="bisacsh">
  <topic>LANGUAGE ARTS &amp; DISCIPLINES / Library &amp; Information Science / Cataloging &amp; Classification</topic>
</subject>
<subject authority="gnd">
  <topic>XML</topic>
</subject>

<relatedItem type="series" displayLabel="Series">
  <titleInfo>
    <title>Third millennium cataloging</title>
  </titleInfo>
</relatedItem>

<note type="bibliography">Includes bibliographical references and index.</note>

<identifier type="isbn">9781598845198 (pbk. : acid-free paper)</identifier>
<identifier type="isbn">1598845195 (pbk. : acid-free paper)</identifier>
<identifier type="isbn">9781610692915 (ebook)</identifier>
<identifier type="lccn">2013006978</identifier>

<location>
  <physicalLocation>University of Illinois at Urbana-Champaign. Library</physicalLocation>
  <copyInformation>
    <subLocation>Social Sciences, Health, and Education</subLocation>
    <shelfLocator>Z678.93.X54 C65 2013</shelfLocator>
    <note>Available</note>
  </copyInformation>
</location>
<itemIdentifier type="copyNumber">1</itemIdentifier>
</copyInformation>
</holdingSimple>
</location>
</mods>
Valid XML Exercise
Solution for Resource #4

Available online
http://quest.library.illinois.edu/ALA2015/ALCTS-ExerciseSolutions/Exercise2/Valid/Sample4-MODS.xml

<?xml version="1.0" encoding="utf-8"?>
<titleInfo>
<title>XML in libraries</title>
</titleInfo>

<name type="personal">
<namPart type="family">Tennant</namPart>
<namPart type="given">Roy</namPart>
<Role>
<roleTerm type="text" authority="marcrelator">editor</roleTerm>
<roleTerm type="code" authority="marcrelator">edt</roleTerm>
</Role>
</name>

<typeOfResource>text</typeOfResource>

<originInfo>
<place>
<placeTerm type="code" authority="marccountry">nyu</placeTerm>
<placeTerm type="text">New York</placeTerm>
</place>
<publisher>Neal-Schuman Publishers</publisher>
<copyrightDate>2002</copyrightDate>
<issuance>monographic</issuance>
</originInfo>

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

<physicalDescription>
<form authority="marcform">print</form>
<extent>xi, 213 p. : ill. ; 23 cm.</extent>
</physicalDescription>

<tableOfContents>
Updating MARC records with XMLMARC / Kevin S. Clarke -- Searching and retrieving XML records via the Web / Theo van Veen -- Improving interlibrary loan with XML / Kyle Banerjee -- Harnessing Oracle and XT for finding aid dissemination and search / Leslie Myrick -- Creating a unified e-government portal using XML / Lloyd Sokvitne and Jan Lavelle -- Expediting the work of the indexer with XML / Walter Lewis, Gail Richardson, and Geoff Cannon -- Using XML to federate collections : the Legacy Tobacco Documents Library / Heidi Schmidt -- Publishing books online at eScholarship / Roy Tennant -- Building XML databases with Zope and Castor / Art Rhyno -- Migrating native law cases from HTML to XML / Darlene Fichter -- Transforming word processing documents into XML : electronic scholarly publishing at the University of Michigan / Brian Rosenblum -- Encoding digital objects with METS / Jerome McDonough -- Integrating systems with XML-based Web services / Don Gourley.
</tableOfContents>

<note type="bibliography">Includes bibliographical references and index.</note>
<?xml version="1.0" encoding="utf-8"?>
<titleInfo>
<title>50-50: Fighting Chicago's Crime Trusts</title>
</titleInfo>
<name type="personal">
<namePart type="family">French</namePart>
<namePart type="given">George W.</namePart>
</name>
<typeOfResource>image</typeOfResource>
<accessCondition>Images in this collection were digitized through the University of Illinois Library's participation in the Open Content Alliance and may be used freely. Attribution to the University of Illinois is appreciated. High-resolution images can be downloaded from the Internet Archive at www.archive.org. For further information, contact dcc@library.uiuc.edu.</accessCondition>
<originInfo>
<place>
<placeTerm type="code" authority="marccountry">ilu</placeTerm>
<placeTerm type="text" authority="marccountry">Illinois</placeTerm>
</place>
<dateCreated>1916</dateCreated>
</originInfo>
<physicalDescription>
<internetMediaType>image/jp2</internetMediaType>
</physicalDescription>
<language>
<languageTerm>English</languageTerm>
</language>
<note>Cover of the book "50-50": Fighting Chicago's Crime Trusts.</note>
<subject>
<topic>Book covers</topic>
</subject>
<subject>
<topic>Covers (Illustration)</topic>
</subject>
<subject>
<geographic>Chicago (Ill.)</geographic>
</subject>
<subject>
<temporal>1910-1019</temporal>
</subject>
<relatedItem type="isPartOf" displayLabel="Collection">
<titleInfo>
<title>Picture Chicago</title>
</titleInfo>
</relatedItem>
</mods>
<title>50-50: fighting Chicago's crime trusts</title>

<name>
  <namePart type="family">Chamberlin</namePart>
  <namePart type="given">Henry Barrett</namePart>
  <namePart type="date">b. 1867.</namePart>
</name>

<titleInfo>
  <title>50-50: fighting Chicago's crime trusts</title>
</titleInfo>

<identifier type="local">5050fightingchic00cham_09.jp2</identifier>
<identifier type="uri">http://imagesearchnew.library.illinois.edu/cdm/ref/collection/chicago/id/17</identifier>

<relatedItem type="host" displayLabel="Contained in the Book">
  <titleInfo>
    <title>50-50: fighting Chicago's crime trusts</title>
  </titleInfo>
  <name>
    <namePart type="family">Chamberlin</namePart>
    <namePart type="given">Henry Barrett</namePart>
    <namePart type="date">b. 1867.</namePart>
  </name>
  <location displayLabel="Full text">
    <url>http://www.archive.org/stream/5050fightingchic00cham#page/n2/mode/1up</url>
  </location>
</relatedItem>
Valid XML Exercise
An **INCORRECT** Solution for Resource #4

Available online

[http://quest.library.illinois.edu/ALA2015/ALCTS-ExerciseSolutions/Exercise2/InValid/Sample4-MODS-InValid.xml](http://quest.library.illinois.edu/ALA2015/ALCTS-ExerciseSolutions/Exercise2/InValid/Sample4-MODS-InValid.xml)

```xml
<?xml version="1.0" encoding="utf-8"?>
  <titleInfo>
    <title>XML in libraries</title>
  </titleInfo>
  <name type="personal">
    <namePart type="family">Tennant</namePart>
    <namePart type="given">Roy</namePart>
    <role>
      <roleTerm type="text" authority="marcrelator">editor</roleTerm>
      <roleTerm type="code" authority="marcrelator">edt</roleTerm>
    </role>
  </name>
  <copyrightDate>2002</copyrightDate>
  <originInfo>
    <placeTerm type="code" authority="marccountry">nyu</placeTerm>
    <placeTerm type="text">New York</placeTerm>
    <publisher>Neal-Schuman Publishers</publisher>
    <copyrightDate>2002</copyrightDate>
    <issuance>monographic</issuance>
  </originInfo>
  <language>
    <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
  </language>
  <physicalDescription>
    <form authority="marcform">print</form>
    <extent>xi, 213 p. : ill. ; 23 cm.</extent>
  </physicalDescription>
  <tableOfContents>Updating MARC records with XMLMARC / Kevin S. Clarke -- Searching and retrieving XML records via the Web / Theo van Veen -- Improving interlibrary loan with XML / Kyle Banerjee -- Harnessing Oracle and XT for finding aid dissemination and search / Leslie Myrick -- Creating a unified e-government portal using XML / Lloyd Sokvitne and Jan Lavelle -- Expediting the work of the indexer with XML / Walter Lewis, Gail Richardson, and Geoff Cannon -- Using XML to federate collections : the Legacy Tobacco Documents Library / Heidi Schmidt -- Publishing books online at eScholarship / Roy Tennant -- Building XML databases with Zope and Castor / Art Rhyno -- Migrating native law cases from HTML to XML / Darlene Fichter -- Transforming word processing documents into XML : electronic scholarly publishing at the University of Michigan / Brian Rosenblum -- Encoding digital objects with METS / Jerome McDonough -- Integrating systems with XML-based Web services / Don Gourley.</tableOfContents>
  <note type="bibliography">Includes bibliographical references and index.</note>
  <subject authority="lcsh">
    <topic>XML (Document markup language)</topic>
  </subject>
  <subject authority="lcsh">
    <topic>Libraries</topic>
  </subject>
  <subject authority="gtt">
    <topic>XML</topic>
  </subject>
</mods>
```