Learning Module 4. Representing Metadata in XML Syntax
Week 4 Presentation
XML: Extensible Markup Language

• A W3C Recommendation” = standard: http://www.w3.org/TR/xml/

• Developed in late 1990s specifically for the web environment

• A way to describe data for purposes of information exchange
  – **What the data is** (semantic markup) as opposed to **how the data should be displayed** (formatting markup in pre-HTML5 HTML)

• Provides the framework and specifications – similar with HTML conventions – for marking up data
  – e.g., angle brackets "< >" to enclose tags that indicate elements: "<...>" to “open” the element, "</...>" to “close” the element
<table>
<thead>
<tr>
<th><strong>Object Name</strong></th>
<th>Dress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object ID</strong></td>
<td>1956.274.022</td>
</tr>
<tr>
<td><strong>Collection</strong></td>
<td>Costume</td>
</tr>
<tr>
<td><strong>Grand Parent</strong></td>
<td>3: Personal Artifacts</td>
</tr>
<tr>
<td><strong>Parent</strong></td>
<td>Clothing -- Outerwear</td>
</tr>
<tr>
<td><strong>Early Date</strong></td>
<td>1912</td>
</tr>
<tr>
<td><strong>Late Date</strong></td>
<td>1915</td>
</tr>
<tr>
<td><strong>Received From</strong></td>
<td>Roehm, Dorothy M.</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>Dresses</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Women's black silk velvet, one piece dress having short sleeves, a V-neck, and an ankle length skirt with a peg top. The bodice has black net over white net, and sleeves of plain, double-thick black net. The bodice is trimmed with black and crystal sequins, and jet and crystal beads. The sides of the dress have triangular panniers, shaped with weights. Worn by Mrs. Wm. Traub.</td>
</tr>
</tbody>
</table>
<objectName>Dress</objectName>
<objectID>1956.274.022</objectID>
collection>Costume</collection>
<grandParent>3: Personal Artifacts</grandParent>
<parent>Clothing -- Outerwear</parent>
<earlyDate>1912</earlyDate>
<lateDate>1915</lateDate>
<receivedFrom>
<person>Roehm, Dorothy M.</person>
</receivedFrom>
<subject>Dresses</subject>
<description>Women's black silk velvet, one piece dress having short sleeves, a V-neck, and an ankle length skirt with a peg top. The bodice has black net over white net, and sleeves of plain, double-thick black net. The bodice is trimmed with black and crystal sequins, and jet and crystal beads. The sides of the dress have triangular panniers, shaped with weights. Worn by Mrs. Wm. Traub.</description>
XML

Well-formedness criteria:

- Closing tags
- Case sensitivity (same case in start & end tags for the element)
- Proper nesting of elements

This record is well-formed

```
<objectName>Dress</objectName>
<objectID>1956.274.022</objectID>
<collection>Costume</collection>
<grandParent>3: Personal Artifacts</grandParent>
<parent>Clothing -- Outerwear</parent>
<earlyDate>1912</earlyDate>
<lateDate>1915</lateDate>
<receivedFrom>
  <person>Roehm, Dorothy M.</person>
</receivedFrom>
<subject>Dresses</subject>
<description>Women's black silk velvet, one piece dress having short sleeves, a V-neck, and an ankle length skirt with a peg top. The bodice has black net over white net, and sleeves of plain, double-thick black net. The bodice is trimmed with black and crystal sequins, and jet and crystal beads. The sides of the dress have triangular panniers, shaped with weights. Worn by Mrs. Wm. Traub.</description>
```
XML

Well-formedness criteria:
Closing tags
Case sensitivity (start & end tags)
Proper nesting of elements

How about this record?

<record2>
  <creator>Lennon, John, 1940-1980</creator>
  <address>McCartney, Paul</address>
  <Title>Let it be</Title>
  <type>Song (music)</type>
  <format>Vinyl record 7”</format>
  <date>1970</date>
  <subject>Mothers Love Rock music, 1961-1970</subject>
</record2>
Well-formedness criteria:

- Closing tags
- Case sensitivity (start & end tags)
- Proper nesting of elements

How about this record? NO
Metadata Record Example: basic XML

Well-formedness criteria:

• Closing tags
• Case sensitivity (start & end tags)
• Proper nesting of elements

How about this record? NO

---

<record2>
  <creator>Lennon, John, 1940-1980</creator>
  <address>McCartney, Paul</address>
  <Title>Let it be</Title>
  <type>Song (music)</type>
  <format>Vinyl record 7”</format>
  <date>1970</date>
  <subject>Mother Love Rock music, 1961-1970</subject>
</record2>
Well-formedness criteria:

• Closing tags
• Case sensitivity (start & end tags)
• Proper **nesting** of elements

How about this record?  NO
Other problems with this record: use of wrong element to hold data

Metadata Record Example: basic XML

<record2>
  <creator>Lennon, John, 1940-1980</creator>
  <address>McCartney, Paul</address>
  <Title>Let it be</Title>
  <type>Song (music)</type>
  <format>Vinyl record 7”</format>
  <date>1970</date>
  <subject>Mothers Love Rock music, 1961-1970</subject>
</record2>

<subject/>

Other problems with this record:
use of wrong element to hold data
<?xml version="1.0"?>
<!DOCTYPE metadata record SYSTEM "metadataRecord.dtd">
[OR xmlns:dhmc="http://detroithistorical.org/detroit-historical-museum/costume-metadata"]

<! --Edited with MS Powerpoint-->
<record1>
  <objectName>Dress</objectName>
  <objectID>1956.274.022</objectID>
  <collection>Costume</collection>
  <grandParent>3: Personal Artifacts</grandParent>
  <parent>Clothing -- Outerwear</parent>
  <earlyDate>1912</earlyDate>
  <lateDate>1915</lateDate>
  <receivedFrom>
    <person>Roehm, Dorothy M.</person>
  </receivedFrom>
  <subject>Dresses</subject>
  <description>Women’s black silk velvet, one piece dress having short sleeves, a V-neck, and an ankle length skirt with a peg top. The bodice has black net over white net, and sleeves of plain, double-thick black net. The bodice is trimmed with black and crystal sequins, and jet and crystal beads. The sides of the dress have triangular panniers, shaped with weights. Worn by Mrs. Wm. Traub.</description>
</record1>
<?xml version="1.0"?>
<!DOCTYPE metadata record SYSTEM "metadataRecord.dtd">

[OR <xmlns:dhmc="http://detroithistorical.org/detroit-historical-museum/costume-metadata">]

<! --Edited with MS Powerpoint-->

<record1>
  <objectName>Dress</objectName>
  <objectID>1956.274.022</objectID>
  <collection>Costume</collection>
  <grandParent>3: Personal Artifacts</grandParent>
  <parent>Clothing -- Outerwear</parent>
  <earlyDate>1912</earlyDate>
  <lateDate>1915</lateDate>
  <receivedFrom>
    <person>Roehm, Dorothy M.</person>
  </receivedFrom>
  <subject>Dresses</subject>
  <description>Women's black silk velvet, one piece dress having short sleeves, a V-neck, and an ankle length skirt with a peg top. The bodice has black net over white net, and sleeves of plain, double-thick black net. The bodice is trimmed with black and crystal sequins, and jet and crystal beads. The sides of the dress have triangular panniers, shaped with weights. Worn by Mrs. Wm. Traub.</description>
</record1>
<?xml version="1.0"?>
<xmlns:dc="http://purl.org/dc/elements/1.1">
<! --Edited with MS Powerpoint-->
<record>
  <creator>Shevchenko, Taras, 1814-1861</creator>
  <title>The Gypsy Fortune Teller</title>
  <title>Циганка-ворожка</title>
  <type>Painting</type>
  <format>Watercolor on Bristol paper</format>
  <format>26.5 X 20.7 cm</format>
  <date>1841</date>
  <coverage>Ukraine</coverage>
  <subject>Fortune-telling</subject>
  <subject>Women</subject>
  <subject>Children</subject>
  <subject>Dogs</subject>
  <relation>Zavalina, O., 2016. ALA eCourse LM4</relation>
</record>
Creating **DC Simple** record in XML in this course

In the text editor, start your record in XML syntax with the template looking something like this:

```xml
<?xml version="1.0" encoding="UTF-8"?>
    <creator>[...]</creator>
    <contributor>[...]</contributor>
    <coverage>[...]</coverage>
    <date>[...]</date>
    <description>[...]</description>
    <format>[...]</format>
    <identifier>[...]</identifier>
    <language>[...]</language>
    <publisher>[...]</publisher>
    <relation>[...]</relation>
    <rights>[...]</rights>
    <source>[...]</source>
    <subject>[...]</subject>
    <title>[...]</title>
    <type>[...]</type>
</dc>
```

Then replace ellipses ([...]) with data values (a.k.a. content) from your Exercise 2 & repeat the applicable elements as needed. For elements that are not applicable, use “n/a” as data value.
<table>
<thead>
<tr>
<th><strong>Element</strong></th>
<th><strong>Data value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
<td>Trends in metadata practices: A longitudinal study of collection federation</td>
</tr>
<tr>
<td><strong>Creator</strong></td>
<td>Palmer, Carole L.</td>
</tr>
<tr>
<td><strong>Creator</strong></td>
<td>Zavalina, Oksana L.</td>
</tr>
<tr>
<td><strong>Creator</strong></td>
<td>Mustafoff, Megan</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Presents results from a longitudinal analysis of IMLS DCC development trends between 2003 and 2006. Changes in metadata applications have not been pronounced. However, multi-scheme use has become less common, and use of Dublin Core remains high, even as recognition of its limitations grows</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>Descriptive metadata</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>Metadata schemes</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>Interoperability</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td>USA</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td>2003-2006</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2007</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>pdf</td>
</tr>
<tr>
<td><strong>Identifier</strong></td>
<td><a href="http://dl.acm.org/citation.cfm?doid=1255175.1255251">http://dl.acm.org/citation.cfm?doid=1255175.1255251</a></td>
</tr>
<tr>
<td><strong>Identifier</strong></td>
<td>10.1145/1255175.1255251</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>English</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>Association for Computing Machinery</td>
</tr>
<tr>
<td><strong>Relation</strong></td>
<td>Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries</td>
</tr>
<tr>
<td><strong>Relation</strong></td>
<td>9781595936448</td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>Text is copyrighted ACM/IEEE-CS and original authors</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>conference proceedings paper</td>
</tr>
</tbody>
</table>

**NOTE:** *Source* and *Contributor* elements are not used (n/a to this information object)
<?xml version="1.0" encoding="UTF-8"?>
  <title>Trends in metadata practices: A longitudinal study of collection federation</title>
  <creator>Palmer, Carole L.</creator>
  <creator>Zavalina, Oksana L.</creator>
  <creator>Mustafoff, Megan</creator>
  <description>Presents results from a longitudinal analysis of IMLS DCC development trends between 2003 and 2006. Changes in metadata applications have not been pronounced. However, multi-scheme use has become less common, and use of Dublin Core remains high, even as recognition of its limitations grows</description>
  <subject>Descriptive metadata</subject>
  <subject>Metadata schemes</subject>
  <subject>Interoperability</subject>
  <coverage>USA</coverage>
  <coverage>2003-2006</coverage>
  <date>2007</date>
  <format>pdf</format>
  <identifier>http://dl.acm.org/citation.cfm?doid=1255175.1255251</identifier>
  <identifier>10.1145/1255175.1255251</identifier>
  <language>English</language>
  <publisher>Association for Computing Machinery</publisher>
  <relation>Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries</relation>
  <relation>9781595936448</relation>
  <rights>Text is copyrighted ACM/IEEE-CS and original authors</rights>
  <type>conference proceedings paper</type>
</dc>
Example of MODS record for the same paper (from LM3) in XML syntax (p.1-2)

<?xml version="1.0"?
<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"
version="3.6"
xsi:schemaLocation="http://www.loc.gov/standards/mods/v3/mods-3-6.xsd">
<titleInfo>
<title>Trends in metadata practices </title>
<subTitle>A longitudinal study of collection federation</subTitle>
</titleInfo>
<name
authority="naf"
authorityURI="http://id.loc.gov/authorities/names/"
type="personal">
<namePart type="family">Palmer,</namePart>
<namePart type="given">Carole L. ,</namePart>
<role><roleTerm
authority="marcrelator" type="term">author </roleTerm></role>
<affiliation>University of Illinois </affiliation>
</name>
<name
authority="naf"
authorityURI="http://id.loc.gov/authorities/names/"
type="personal">
<namePart type="family">Zavalina </namePart>
<namePart type="given">Oksana </namePart>
<role><roleTerm
authority="marcrelator" type="term">author </roleTerm></role>
<affiliation>University of Illinois </affiliation>
</name>
<typeOfResource>text</typeOfResource>
<genre authority="marcgt">conference publication</genre>
<originInfo><place><placeTerm authority="marccountry" type="code">nyu</placeTerm></place>
<publisher>Association for Computing Machinery</publisher>
<copyrightDate>2007</copyrightDate>
</originInfo>
<language><languageTerm authority= "marclanguage" type="code">eng</languageTerm></language>
<physicalDescription><internetMediaType>application/pdf</internetMediaType>
<extent unit="page">10</extent>
digitalOrigin>born digital</digitalOrigin></physicalDescription>
<abstract>Presents results from a longitudinal analysis of IMLS DCC development trends between 2003 and 2006. Changes in metadata applications have not been pronounced. However, multi-scheme use has become less common, and use of Dublin Core remains high, even as recognition of its limitations grows.</abstract>
<targetAudience authority="marctarget"> specialized</targetAudience>
<subject authority="lcsh" authorityURI="http://id.loc.gov/authorities/subjects/"><topic>Metadata</topic></subject>
</mods>
Example of MODS record for the same paper (from LM3) in XML syntax (p.3-5)

```xml
<subject><topic>Descriptive metadata</topic></subject>
<subject><geographic>United States</geographic></subject>
<subject><temporal>2003-2006</temporal></subject>
<subject><topic>Metadata schemes</topic></subject>
<subject><topic>Interoperability</topic></subject>
<classification authority="lcc">Z666.7</classification>

<relatedItem type="host" displayLabel="Published in:"
   <name type="personal" authority="naf" authorityURI="http://id.loc.gov/authorities/names"><namePart>Rasmussen, Edie M.</namePart><role><roleTerm authority="marcrelator" type="code">edt</roleTerm></role></name>
   <name type="personal" authority="naf" authorityURI="http://id.loc.gov/authorities/names"><namePart>Larson</namePart><namePart type="given">Ray R.</namePart><role><roleTerm authority="marcrelator" type="code">edt</roleTerm></role></name>
   <name type="personal" authority="naf" authorityURI="http://id.loc.gov/authorities/names"><namePart>Sugimoto</namePart><namePart type="given">Shigeo</namePart><namePart type="date">1953-</namePart><role><roleTerm authority="marcrelator" type="code">edt</roleTerm></role></name>
   <identifier type="isbn">9781595936448</identifier>
</relatedItem>

<identifier type="doi">doi:10.1145/1255175.1255251</identifier>
<location><url>http://dl.acm.org/citation.cfm?doid=1255175.1255251</url></location>
<accessCondition type="use and reproduction">Subscription to ACM Digital Library</accessCondition>
<recordInfo>
   <recordCreationDate encoding="w3cdtf">2016-07-01</recordCreationDate>
</recordInfo>

Dr. Oksana L. Zavalina, © 2015-2016
Creating MODS XML documents in this course with text editor (e.g., Sublime)

1. Start with a template based on your Exercise 3 record structure: elements, subelements, and the number of instances for each; element attributes, and attribute values.

2. Populate the template with data values.

3. For non-applicable elements, enter “[n/a]”.

The document with this sample XML template is posted for you in course materials.
MODS record is always more detailed than DC record for the same object: should contain everything DC record contains and more!

```xml
<?xml version="1.0"?>
<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" version="3.6"
     xsi:schemaLocation="http://www.loc.gov/standards/mods/v3/mods-3-6.xsd">

<titleInfo>
<title>The gypsy fortune teller</title>
</titleInfo>

<titleInfo>
<title type="translated" lang="ukr">Циганка-ворожка</title>
</titleInfo>

</mods>
```
<name type="personal" authority="naf">
  <namePart type="family">Shevchenko</namePart>
  <namePart type="given">Taras</namePart>
  <namePart type="dates">1814-1861</namePart>
  <role><roleTerm authority="marcrelator" type="text">Author</roleTerm></role>
</name>
<typeOfResource>still image</typeOfResource>

<genre authority="marcgt">picture</genre>

<originInfo>
  <place>Saint Petersburg, Russia</place>
  <dateCreated>1841</dateCreated>
</originInfo>

<physicalDescription>
  <form type="material">watercolor</form>
  <form type="material">Bristol paper</form>
  <form type="technique" authority="aat">aquarelle</form>
  <extent>26.5 X 20.7 cm</extent>
</physicalDescription>
<abstract>This painting shows an episode of fortune-telling at a Ukrainian village: the client is a Ukrainian girl in the national costume, and the fortune teller is a gypsy woman. Both stand outside under the green maple tree next to the gate. Also present in the painting are a child carried by a fortune-teller and a dog.</abstract>

<targetAudience>General</targetAudience>

<note>Created by Shevchenko when he was residing in Saint-Petersburg and studying at the Imperial Academy of Arts. Shevchenko was awarded his third Silver Medal for this painting.</note>
<subject>
    <geographic authority="tgn">Ukraina</geographic>
</subject>

<subject>
    <topic authority="lcsh">Fortune-telling</topic>
</subject>

<subject>
    <topic authority="lcsh">Women</topic>
</subject>

<subject>
    <topic authority="lcsh">Children</topic>
</subject>

<subject>
    <topic authority="lcsh">Dogs</topic>
</subject>

<classification authority="ddc">133.3</classification>
<location>
  <physicalLocation>
    Taras Shevchenko National Museum
  </physicalLocation>
</location>

<accessCondition type="use and reproduction">
  Use of this public-domain resource is unrestricted.
</accessCondition>
<relatedItem type="isReferencedBy">
  <titleInfo>
    <title>ALA eCourse Metadata and Linked Data</title>
  </titleInfo>
  <name type="personal" authority="lcnaf">
    <namePart type="family">Zavalina</namePart>
    <namePart type="given">Oksana</namePart>
    <role>
      <roleTerm authority="marcrelator" type="text">Author</roleTerm>
    </role>
  </name>
  <originInfo>
    <dateCreated>2015</dateCreated>
    <location><url>https://example.url</url></location>
  </originInfo>
</relatedItem>
<recordInfo>
  <recordCreationDate encoding="w3cdtf">2016-01-01</recordCreationDate>
  <recordChangeDate encoding="w3cdtf">2016-06-06</recordChangeDate>
  <languageOfCataloging>
    <languageTerm authority="iso639-2">eng</languageTerm>
  </languageOfCataloging>
</recordInfo>
MARC XML metadata

https://www.loc.gov/standards/marcxml/

- **MARC 21 compatible without any loss**
  
  - Same element set as in MARC
  
  - Same element names (e.g., **245**) and same subelement names (e.g., **264 _1 $c**) as MARC
  
  - Same semantics as MARC

- **XML syntax instead of MARC syntax**
  
  - Encoded in **XML** syntax for reuse of library data in the broader web environment

- Used with cataloging rules
  
  - AACR2r or RDA

Dr. Oksana L. Zavalina, © 2015-2016
<record>
  <leader>01301cam a2200277 a 450 </leader>
  <controlfield tag="001">15258260 005 20100707091951.0</controlfield>
  <controlfield tag="008">080411s2008 nyua b 001 0 eng</controlfield>
  <datafield tag="010" ind1="" ind2="" subfield code="a">2008015176</datafield>
  <datafield tag="020" ind1="" ind2="" subfield code="a">9781555706357</datafield>
  <datafield tag="040" ind1="" ind2="" subfield code="a">DLC</datafield>
  <datafield tag="050" ind1="" ind2="" subfield code="b">Z666.7</datafield>
  <datafield tag="052" ind1="" ind2="" subfield code="b">Z46 2008</datafield>
  <datafield tag="082" ind1="" ind2="" subfield code="b">025.3</datafield>
  <datafield tag="100" ind1="" ind2="" subfield code="d">1956</datafield>
  <datafield tag="245" ind1="" ind2="" subfield code="c">Marcia Lei Zeng and Jian Qin</datafield>
  <datafield tag="260" ind1="" ind2="" subfield code="c">©2008</datafield>
  <datafield tag="300" ind1="" ind2="" subfield code="b">xvii, 365 p.</datafield>
  <datafield tag="504" ind1="" ind2="" subfield code="a">Includes bibliographical references (p. 327-353) and index.</datafield>
  <datafield tag="505" ind1="" ind2="" subfield code="a">Introduction -- Current standards -- Schemas : structure and semantics -- Schemas : syntax -- Metadata records -- Metadata services -- Metadata quality measurement and improvement</datafield>
  <datafield tag="650" ind1="" ind2="" subfield code="d">1956-</datafield>
  <datafield tag="650" ind1="" ind2="" subfield code="d">Information organization</datafield>
  <datafield tag="700" ind1="" ind2="" subfield code="d">Qin, Jian</datafield>
  <datafield tag="856" ind1="" ind2="" subfield code="3">Table of contents</datafield>
  <datafield tag="886" ind1="" ind2="" subfield code="u">http://www.loc.gov/catdir/typq0816/2008015176.html</datafield>
</record>
<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Metadata /</subfield>
  <subfield code="c">Marcia Lei Zeng and Jian Qin.</subfield>
</datafield>
Efficient way of creating MARCXML records in this course with text editor (e.g., Sublime)

1. Start with the following basic template and expand it based on the MARC fields and subfields in the snippet.

2. Replace ellipses with field tags, field indicators, subfield codes and data values found in MARC 21 record.