Exercise 6: Metadata Quality and Interoperability: ANSWER KEY


Exercise Description: This exercise helps develop understanding of metadata quality through evaluation of metadata records and identifying problems with their quality. This exercise also helps develop understanding of mapping between various metadata schemes through application of a crosswalk to convert a metadata record from one metadata scheme to another.

Exercise Instructions:

Part 1. Please examine the following Dublin Core and MODS metadata records (in XML syntax) which describe the same information object that you described in Exercises 2, 3, and 4. The records are located in the left-hand column of the table below. Please enter in the right-hand column of the table your notes about quality of these records using the three major criteria of metadata quality: completeness, accuracy, and consistency.

Please be specific, e.g.:
- if you observed incompleteness in the record, name the metadata elements/subelements that are missing and/or the attributes of the information object being described that are not represented in the record;
- if you are commenting on inaccurate data values, please name the elements with inaccurate data values and specify what the correct data value should be, etc.

Consult course materials (readings and instructor’s videorecorded presentations) as you work.

<table>
<thead>
<tr>
<th>Metadata records to evaluate</th>
<th>Your comments on metadata quality of these records</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECORD 1 (see below):</td>
<td></td>
</tr>
<tr>
<td>&lt;?xml version=&quot;1.0&quot;?&gt;</td>
<td>Comments on completeness of this Dublin Core record:</td>
</tr>
</tbody>
</table>
This record is incomplete as:
  - 2 applicable elements are missing: Publisher and Format
  - Description element has an "n/a" as a data value (the actual data value – summary or table of contents – is missing).

**Comments on accuracy of this Dublin Core record:**

There are multiple issues with accuracy of this record:
  - The Source element has the value that does not belong to this element
  - The 1st instance of the Coverage element has inaccurate value 1999 (this work is not about 1999 so this value should not be used with this element)
  - Data values in Creator and Contributor elements are inaccurate: first and middle name mixed up for Dixit, a typo in the first name for Rai
  - Inaccurate nesting of the Date element: two data values and two end tags but only one start tag
  - The start and end tags for Description element do not match (the end tag is for Abstract element) which affects proper nesting of the element

**Comments on consistency of this Dublin Core record:**

There are multiple issues with consistency of this record:
  - The order of name components in two elements containing personal name -- Creator and Contributor -- is inconsistent (the creator name is inverted while the contributor name is not)
  - Missing namespace ("dc") in the XML start tag for Relation element and end tags for Contributor and Date element
  - The element name is missing in the end tag for Date element

Any other comments on the quality of this Dublin Core record:

The problems with XML nesting mentioned above, as well as with case-sensitivity (the end tags for Creator and Language elements contain upper case characters while the start tags do not), make this XML document non-well-formed and invalid.

**RECORD 2 (see below):**

```xml
<xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<mods version="3.6">
  <mods:titleInfo>
    <mods:title>Adventures of a Nepali frog</mods:title>
    <mods:name type="corporate" authority="lcnaf">
      <mods:namePart>Subhasha Rai</mods:namePart>
    </mods:name>
  </mods:titleInfo>
</mods>
</xmlns:xsi>
```

**Comments on completeness of this MODS record:**

This record is incomplete as:
  - Six applicable top-level elements are missing: Type of Resource, Abstract, Table of Contents, Target Audience, Note, and Classification
  - The instances of Name element for Dixit and Rai are missing
<mods:namePart>Rato Bangala Kitab</mods:namePart>
<mods:role><mods:roleTerm>publisher</mods:roleTerm></mods:role>
</mods:name>
<mods:genre authority="marcgt">juvenile works</mods:genre>
<mods:originInfo>...
<mods:place>n/a</mods:place>
<mods:copyrightDate>1999</mods:copyrightDate>
</mods:originInfo>
<mods:language>
<mods:languageTerm authority="marclanguage">Nepalese</mods:languageTerm>
</mods:language>
<mods:physicalDescription>
<mods:extent>102 pages</mods:extent>
</mods:physicalDescription>
<mods:subject authority="lcsh">
<mods:topic>Frogs</mods:topic>
</mods:subject>
<mods:subject authority="lcnaf">
<mods:geographic>Nepal</mods:geographic>
</mods:subject>
<mods:subject authority="lcsh">
<mods:temporal>2003</mods:temporal>
</mods:subject>
<mods:relatedItem>
<mods:titleInfo>
Internet Children’s Digital Library</mods:titleInfo>
</mods:relatedItem>
<mods:identifier authority="isbn">http://goo.gl/qmOxNk</mods:identifier>

- Applicable subelements of Origin Info element -- Date Issued and Publisher – are missing
- Applicable subelement Internet Media Type of Physical Description element is missing

Comments on accuracy of this MODS record:
There are multiple issues with accuracy of this record:
- The name of organization that published this work and the role "publisher" are incorrectly used in the Name element
- The "juvenile works" data value for Genre element is not from the marcgt controlled vocabulary
- The "n/a" data value for Place element is inaccurate and does not make sense
- The "Kitab" data value for Publisher element is inaccurate (should be "Rato Bangala Kitab")
- The "Nepalese" data value for Language element is incorrect
- The "born digital" data value for Digital Origin element is incorrect
- The "2003" data value for Temporal subelement of Subject element is incorrect

Comments on consistency of this MODS record:
- The nesting of elements is inconsistent (please see below for details)

Any other comments on the quality of this MODS record:
There are multiple issues affecting well-formedness and validity of this XML document:
- Incorrect nesting with Language element (the end tag for subelement Language Term appear after the end tag for parent element Language)
- Missing namespace ("mods:" in the end tag for Geographic element results in a mismatch Between open tag and close tag for the element
- Missing start tag for Title subelement both under Title Info main element and under Related Item
- Missing end tag for Title Info subelement of Related Item element
- Missing start tag for Language Term subelement of the Language of Cataloging element under Record Info
- Improper nesting in Record Info element – the end tag for the element appears before one of the two subelements of this element – Language of Cataloging
Part 2. Please consult this crosswalk table http://www.alastore.ala.org/nealschuman/companionwebsite/metadata-digital-collections/MDC_DC-MODS-VRA_Mapping_Table.pdf (1st and 4th columns for DC and VRA Core 4.0) to manually transform this VRA Core 4.0 record http://core.vraweb.org/examples/html/example026_full_xml.html for the work Autumn – On the Hudson River into the Dublin Core record in RDF/XML syntax (just like the one you created in Exercise 5 Part 1). If the crosswalk table PDF document does not open by following the link above, simply copy and paste the URL into your browser window. Consult course materials (readings and instructor’s videorecorded presentations) as you work. Please enter the resulting Dublin Core record in the table below.

You resulting Dublin Core metadata record in RDF/XML syntax:

<?xml version="1.0" encoding="UTF-8" ?>
    <rdf:Description rdf:about="Autumn – On the Hudson River">
        <dc:creator rdf:resource="http://id.loc.gov/authorities/names/n80026719"/>
        <dc:contributor rdf:resource="http://vocab.getty.edu/ulan/500115983"/>
        <dc:coverage>American</dc:coverage>
        <dc:coverage rdf:resource="http://id.loc.gov/authorities/subjects/sh85062785"/>
        <dc:coverage rdf:resource="http://id.loc.gov/authorities/subjects/sh85091984"/>
        <dc:date>1860</dc:date>
        <dc:description>This monumental view of the Hudson River Valley was painted from memory in the artist’s London</dc:description>
    </rdf:Description>
</rdf:RDF>
studio. Cropsey adopted a high vantage point, looking southeast toward the distant Hudson River and the flank of Storm King Mountain. A small stream leads from the foreground, where three hunters and their dogs gaze into the sunlight. All along the meandering tributary there are signs of man’s peaceful coexistence with nature: a small log cabin, grazing sheep, children playing on a bridge, and cows standing placidly in the water. Here, man neither conquers nor is subservient to nature; both coexist harmoniously. In fact, the landscape is depicted as a ready arena for further agricultural expansion. Cropsey’s painting is a celebration of American nationalism.