Choosing a metadata standard for your digital project
Indiana Digital Library Summit Metadata Working Group
Finalized Summer 2004

DESCRIPTIVE METADATA

Dublin Core (DC)

Type of collection: Groups of materials whose organization is not a significant access point to the collection. Generally these materials will not be organized hierarchically. The collection may be in any format, or a mixture of formats. Collections for which item-level description is planned and none of the other metadata needs described in this document apply.

Appropriate metadata standard: Dublin Core Metadata Element Set 1.1

Example collection: Worthington Memory <http://www.worthingtonmemory.org/>

Resources to consult: Dublin Core Metadata Element Set, Version 1.1: Reference Description <http://www.dublincore.org/documents/dces/>; Dublin Core Usage Guide <http://www.dublincore.org/documents/usageguide/>; DCMI Metadata Terms <http://www.dublincore.org/documents/dcmi-terms/>

Usage notes: Simple Dublin Core, consisting of only the 15 core elements, can be used, although much desired robustness (e.g., distinguishing between roles of the creator of a resource, specifying the name of the controlled vocabulary from which a value is selected) is not possible. Qualified Dublin Core, using the DCMI "Other Elements and Element Refinements" as specified in <http://www.dublincore.org/documents/dcmi-terms/#H3> can be used to slightly increase robustness, although the refinements available may not be appropriate for many uses. Creating local qualifiers for Dublin Core elements is another option; however this practice reduces the interoperability of the metadata created. The emerging Dublin Core Library Application Profile <http://dublincore.org/documents/2002/04/16/library-application-profile/index.shtml>, still in Working Draft status, attempts to address these issues.

Encoded Archival Description (EAD)

Type of collection: Archivally-processed groups of materials whose organization and/or provenance is significant. These materials will generally be hierarchically arranged natural groupings assembled by a collector or creator (e.g., the papers or correspondence of a certain individual) rather than less-tightly related groups of materials assembled by a holding institution. A Finding Aid or inventory may or may not already exist. Materials will often be unpublished. Description at the collection level is necessary, lower levels of description may or may not be appropriate. Materials described may or may not be available in digital form.

Appropriate metadata standard: EAD 2002

Example collection: Finding Aids in the Online Archive of California <http://www.oac.cdlib.org/>

Resources to consult: Official EAD Version 2002 Web Site <http://www.loc.gov/ead/>;
  EAD Help Pages, especially the EAD 2002 Cookbook
  <http://www.iath.virginia.edu/ead/>
Usage notes: EAD 2002 can be used to describe the collection at the item, folder-only, or
  collection-only level. One EAD document should be created for each collection.

Text Encoding Initiative (TEI)

Type of collection: Text collections intended for full-text searching in an online
  environment. Full text may or may not be intended to be used together with page
  images of the original document.
Appropriate metadata standard: TEI P4
Example collection: Indiana University, Wright American Fiction, 1851-1875
  <http://www.letrs.indiana.edu/web/w/wright2/>
Resources to consult: TEI Text Encoding in Libraries Guidelines for Best Encoding
  Practices <http://www.indiana.edu/~letrs/tei/>; TEI Guidelines <http://www.tei-
  c.org/Guidelines2/index.htm/>
Usage notes: Choose an Encoding Level as described in the TEI Text Encoding in
  Libraries Guidelines for Best Encoding Practices. Bibliographic information for
  resource discovery is encoded in the TEI Header. The entire full text of the resource
  is marked up structurally in the bulk of the TEI document, and this markup is used for
  powerful full-text searching.

Metadata Object Description Schema (MODS)

Type of collection: Materials with existing item-level MARC cataloging
Appropriate metadata standard: MODS 3.0
Example collection: University of Chicago Chopin Early Editions Project
  <http://chopin.lib.uchicago.edu/>
Resources to consult: MODS official Web site <http://www.loc.gov/standards/mods/>;
  MODS User Guidelines <http://www.loc.gov/standards/mods/v3/mods-
  userguide.html>
Usage notes: Items for which MARC cataloging already exists can be transformed into
  MODS records for use in digital library applications.

Visual Resources Association (VRA) Core

Type of collection: Art images whose users require in-depth indexing and retrieval using
  expert terms for genre, culture, style, period, etc.
Appropriate metadata standard: VRA Core Categories 3.0
Example collection: Cleveland Museum of Art Collections
  <http://www.clevelandart.org/Explore/> (From each item view, click "More
  Information" to see VRA image metadata.)
Resources to consult: VRA Core version 3 home page
  <http://www.vraweb.org/vracore3.htm>; CC:DA Task Force on VRA Core

Categories Summary Report <http://www.libraries.psu.edu/tas/jca/ccda/docs/tf-vra1.pdf>; Cataloging Cultural Objects <http://www.vraweb.org/CCOweb/>.

Usage notes: VRA Core is more robust than Dublin Core for describing art images and metadata in this format is consequently more powerful but more expensive to create. VRA Core contains both "work" records describing an actual art object, and "image" records describing representations of views of that object (slides, digital images, etc.) held by an institution. Best practice in creating VRA Core records is to populate fields using appropriate controlled vocabularies such as ULAN and TGM, and the rules described in Cataloging Cultural Objects <http://www.vraweb.org/CCOweb/>.

Gateway to Educational Materials (GEM)

Type of collection: Learning objects that serve education communities (pre-school, K-12, higher education, vocational and technical training, and lifelong learning). These materials require classification criteria special to the education community, such as education level of the target audience, pedagogical methodology, and standards alignment.

Appropriate metadata standard: Dublin Core Metadata Element Set 1.1 as extended in GEM profile

Example collection: NASA Space Science Education Resource Directory <http://teachspacescience.org/> (From each item view, click "More Information" to see GEM metadata.)

Resources to consult: GEM 2.0 Elements and Semantics <http://www.geminfo.org/Workbench/GEM2_elements.html>; Indiana Humanities Council Smart Desktop metadata profile <http://www.ihc4u.org/sd_metadata.htm>.

Usage notes: Both GEM metadata element set and SDI metadata profile are based on Qualified Dublin Core. GEM element set and profile includes 7 additional elements with detailed qualifiers. Local institute can create qualifiers or metadata elements to meet special request of the local audience and collection, although this practice reduces the interoperability of the metadata created. However, dropping elements from the GEM element set is usually less a problem. One-on-one assistance for those who want to create GEM metadata is available at <http://www.geminfo.org/decision.html>; emerging SDI Application Profile <http://www.ihc4u.org/sd_metadata.htm>, still in working draft status, suggests best practices with respect to the usage of GEM element set.

OTHER TYPES OF METADATA

Metadata Encoding & Transmission Standard (METS)

<http://www.loc.gov/standards/mets/>. METS is an XML schema allowing users to "wrap" existing descriptive metadata in any format with structural, technical, administrative, preservation, and meta-metadata to create a single metadata object for a resource. METS provides extension schemas with recommended technical metadata for still images, audio, and video.

NISO Metadata for Images in XML (NISO MIX)
<http://www.loc.gov/standards/mix/>. MIX is an XML schema implementation of the NISO draft standard Z39.87-2002*, Data Dictionary – Technical Metadata for Digital Still Images*.

PREservation Metadata Implementation Strategies (**PREMIS**)
<http://www.oclc.org/research/projects/pmwg/>. PREMIS is not a metadata element set, rather it is an initiative to develop a set of core elements for preservation metadata for the purpose of long-term preservation of digital objects. Some preliminary recommendations of this group which include some metadata element proposals may be found in the report *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects* <http://www.oclc.org/research/projects/pmwg/pm_framework.pdf>.

Simple (Unqualified) Dublin Core record

```
<?xml version="1.0" encoding="utf-8"?>
<oai_dc:dcCollection xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
   <oai_dc:dc>
      <dc:title>Dante /</dc:title>
      <dc:creator> Eliot, T. S. (Thomas Stearns), 1888-1965. </dc:creator>
      <dc:type>text</dc:type>
      <dc:publisher>London : Faber &amp; Faber,</dc:publisher>
      <dc:date>1929.</dc:date>
      <dc:language>eng</dc:language>
      <dc:description>BB: I002 retro/c</dc:description>
      <dc:description>1</dc:description>
      <dc:subject>Dante Alighieri, 1265-1321</dc:subject>
   </oai_dc:dc>
</oai_dc:dcCollection>
```

Qualified Dublin Core Record

```
<html>
   <head>
      <meta name="DC:title" content="Dante"/>
      <meta name="DC:creator" content="Eliot, T. S. (Thomas Stearns), 1888-1965."/>
      <meta name="DC:type" scheme="DCTERMS.DCMIType" content="text"/>
      <meta name="DC:publisher" content="London : Faber &amp; Faber,"/>
      <meta name="DCTERMS:issued" scheme="DCTERMS.W3CDTF"
content="1929."/>
      <meta name="DC:language" scheme="DCTERMS.ISO639-2" content="eng"/>
      <meta name="DC.identifier" content="ALF9368BB"/>
      <meta name="DC:description" content="69 p. ; 20 cm."/>
      <meta name="DC:subject" scheme="DCTERMS.LCSH" content="Dante Alighieri,
1265-1321--Criticism and interpretation"/>
   </head>
   <body/>
</html>
```

MARC record

=LDR 00833cam 2200253 a 45 0
=001 ALF9368BB
=008 970909s1929\\\\enk\\\\\\\\\\\000\0\eng\\
=010 \\$a   29028820
=035 \\$a(InU)ALF9368BB
=035 \\$a(OCoLC)ocm01084166
=035 \\$9(BY)AJL2891
=040 \\$aDLC$cTOL$dOCL$dKSU$dHRH
=090 \\$aPQ4390$b.E5 1929
=100 1\$aEliot, T. S.$q(Thomas Stearns),$d1888-1965.
=245 10$aDante /$cby T.S. Eliot.
=260 \\$aLondon :$bFaber & Faber,$c1929.
=300 \\$a69 p. ;$c20 cm.
=440 \0$aPoets on the poets ;$vno. 2
=500 \\$a"First published in MCMXXIX by Faber & Faber ... " -- t.p. verso.
=591 \\$aBB: I002 retro/c
=600 00$aDante Alighieri,$d1265-1321$xCriticism and interpretation.
=998 \\$a4/26/1999$b8/25/2000$c5/20/1999
=596 \\$a1
=926 \\$aB-MAIN$b_RCSTACKS$cPQ4390 .E4$dNORMAL$f1

MARCXML record

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<collection xmlns="http://www.loc.gov/MARC21/slim"
   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">
  <record>
    <leader>00833cam a2200253 a 45 0</leader>
    <controlfield tag="001">ALF9368BB</controlfield>
    <controlfield tag="008">970909s1929 enk 000 0 eng </controlfield>
    <datafield tag="010" ind1=" " ind2=" ">
      <subfield code="a"> 29028820</subfield>
    </datafield>
    <datafield tag="035" ind1=" " ind2=" ">
      <subfield code="a">(InU)ALF9368BB</subfield>
    </datafield>
    <datafield tag="035" ind1=" " ind2=" ">
      <subfield code="a">(OCoLC)ocm01084166</subfield>
    </datafield>
    <datafield tag="035" ind1=" " ind2=" ">
      <subfield code="9">(BY)AJL2891</subfield>
    </datafield>
    <datafield tag="040" ind1=" " ind2=" ">
      <subfield code="a">DLC</subfield>
      <subfield code="c">TOL</subfield>
      <subfield code="d">OCL</subfield>
      <subfield code="d">KSU</subfield>
      <subfield code="d">HRH</subfield>
    </datafield>
    <datafield tag="090" ind1=" " ind2=" ">
      <subfield code="a">PQ4390</subfield>
      <subfield code="b">.E5 1929</subfield>
    </datafield>
    <datafield tag="100" ind1="1" ind2=" ">
      <subfield code="a">Eliot, T. S.</subfield>
      <subfield code="q">(Thomas Stearns),</subfield>
      <subfield code="d">1888-1965.</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="0">
      <subfield code="a">Dante /</subfield>
      <subfield code="c">by T.S. Eliot.</subfield>
    </datafield>
    <datafield tag="260" ind1=" " ind2=" ">
      <subfield code="a">London :</subfield>
      <subfield code="b">Faber &amp; Faber,</subfield>
```

```xml
          <subfield code="c">1929.</subfield>
      </datafield>
      <datafield tag="300" ind1=" " ind2=" ">
          <subfield code="a">69 p. ;</subfield>
          <subfield code="c">20 cm.</subfield>
      </datafield>
      <datafield tag="440" ind1=" " ind2="0">
          <subfield code="a">Poets on the poets ;</subfield>
          <subfield code="v">no. 2</subfield>
      </datafield>
      <datafield tag="500" ind1=" " ind2=" ">
          <subfield code="a">&quot;First published in MCMXXIX by Faber &amp; Faber
... &quot; -- t.p. verso.</subfield>
      </datafield>
      <datafield tag="591" ind1=" " ind2=" ">
          <subfield code="a">BB: I002 retro/c</subfield>
      </datafield>
      <datafield tag="600" ind1="0" ind2="0">
          <subfield code="a">Dante Alighieri,</subfield>
          <subfield code="d">1265-1321</subfield>
          <subfield code="x">Criticism and interpretation.</subfield>
      </datafield>
      <datafield tag="998" ind1=" " ind2=" ">
          <subfield code="a">4/26/1999</subfield>
          <subfield code="b">8/25/2000</subfield>
          <subfield code="c">5/20/1999</subfield>
      </datafield>
      <datafield tag="596" ind1=" " ind2=" ">
          <subfield code="a">1</subfield>
      </datafield>
      <datafield tag="926" ind1=" " ind2=" ">
          <subfield code="a">B-MAIN</subfield>
          <subfield code="b">_RCSTACKS</subfield>
          <subfield code="c">PQ4390 .E4</subfield>
          <subfield code="d">NORMAL</subfield>
          <subfield code="f">1</subfield>
      </datafield>
    </record>
</collection>
```

MODS record

```xml
<?xml version="1.0" encoding="UTF-8"?>
<modsCollection xmlns="http://www.loc.gov/mods/v3"
   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-3-0.xsd">
   <mods version="3.0">
      <titleInfo>
         <title>Dante</title>
      </titleInfo>
      <name type="personal">
         <namePart>Eliot, T. S. (Thomas Stearns)</namePart>
         <namePart type="date">1888-1965</namePart>
         <role>
            <roleTerm authority="marcrelator" type="text">creator</roleTerm>
         </role>
      </name>
      <typeOfResource>text</typeOfResource>
      <originInfo>
         <place>
            <placeTerm type="code" authority="marccountry">enk</placeTerm>
         </place>
         <place>
            <placeTerm type="text">London</placeTerm>
         </place>
         <publisher>Faber &amp; Faber</publisher>
         <dateIssued>1929</dateIssued>
         <issuance>monographic</issuance>
      </originInfo>
      <language>
         <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
      </language>
      <physicalDescription>
         <form authority="marcform">print</form>
         <extent>69 p. ; 20 cm.</extent>
      </physicalDescription>
      <note type="statement of responsibility">by T.S. Eliot.</note>
      <note>"First published in MCMXXIX by Faber &amp; Faber ... " -- t.p.
verso.</note>
      <subject authority="lcsh">
         <name type="personal">
            <namePart>Dante Alighieri</namePart>
            <namePart type="date">1265-1321</namePart>
         </name>
         <topic>Criticism and interpretation</topic>
```

```xml
          </subject>
          <relatedItem type="series">
             <titleInfo>
                <title>Poets on the poets ; no. 2</title>
             </titleInfo>
          </relatedItem>
          <identifier type="lccn">29028820</identifier>
          <recordInfo>
             <recordContentSource authority="marcorg">DLC</recordContentSource>
             <recordCreationDate encoding="marc">970909</recordCreationDate>
             <recordIdentifier>ALF9368BB</recordIdentifier>
          </recordInfo>
       </mods>
    </modsCollection>
```

A comparison of five descriptive metadata formats

	DC	QDC	MARC	MARCXML	MODS
Record format	XML RDF (X)HTML	XML RDF (X)HTML	ISO 2709 [ANSI Z39.2]	XML	XML
Field labels	Text	Text	Numeric	Numeric	Text
Reliance on AACR	None	None	Strong	Strong	Implied
Common method of creation	By novices, by specialists, and by derivation	By novices, by specialists, and by derivation	By specialists	By derivation	By specialists and by derivation