Report of the Metadata Working Group

Brian Surratt
for TDL Techs
12/12/05

Agenda

- Review of charge
- MODS profile for ETDs
  - Top level elements
  - Outstanding issues
  - Implementation
- Future projects
  - METS for compound ETDs
  - Rights Management Metadata for ETDs
- Discussion

Review of charge

- Chair: Brian Surratt
- Members: Alisha Little, Ann Mitchell, Jason Thomale, Mary Dabney Wilson, Melinda Flannery, Hillary Spiller
- Reports to: Steering Committee
- Reporting schedule: Monthly written reports via Sharepoint
- Charge: Develop standards based metadata schemas for TDL content.

- Plan of work: First project is to develop a descriptive metadata standard for TDL's first collection.
- Timeline: TDL ETD standard by Dec. 15th.

MODS profile for ETDs

- Why MODS?
  - XML based, web friendly, transportable, processible, configurable, sufficiently descriptive without being too complex, extensible
  - Benefits over MARC: MARC isn't XML based and can't easily be output from web forms. Requires special "cataloging" knowledge and systems to implement
  - Benefits over Dublin Core: DC doesn't have sufficient specificity. DC doesn't specify a syntax and is inconsistently applied. DC isn't extensible

Current draft

1. **Title Information**
2. **Name of Author**
3. **Name of Thesis Advisor**
4. **Names of Committee Members (optional)**
5. **Name of Degree Grantor**
6. **Type of Resource**
7. **Genre**
8. **Origin Information**
9. **Language**
10. **Physical Description**
11. **Abstract**
12. **Subject**
13. **Identifier**
14. **Location**
15. **Degree Information**
16. **Record Information**

Root element

```
<?xml version="1.0" encoding="UTF-8"?>

<mods:mods

xmlns:mods="http://www.loc.gov/mods/v3"
xmlns:etd="http://www.ndltd.org/standards/metadata/etdms/1.0/"
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-3-1.xsd
http://www.ndltd.org/standards/metadata/etdms/1.0/
http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd">
```

Title Information

Mandatory practice: Encode the title information in a <mods:titleInfo> wrapper element. Encode the title proper in a <mods:title> subelement. Encode the subtitle in a <mods:subTitle> subelement.

Optional practice: Other valid subelements or attributes within the <mods:titleInfo> element may be used.

Example:

<mods:titleInfo>
<mods:title>Critical processes and performance measures for patient safety systems in healthcare institutions
</mods:title>
<mods:subTitle>a Delphi study</mods:subTitle>
</mods:titleInfo>

Name of Author

Mandatory practice: Encode information about the name of the author in the <mods:name> wrapper element with the type attribute set to “personal.” Encode the MARC relator term “Author” in the <mods:roleTerm> subelement under the <mods:role> subelement. Encode the various parts of the name in the <mods:namePart> subelement. Include the type attribute in each <mods:namePart> subelement. The “given” and “family” name types are mandatory.

Optional practice: Encode the birthdate in a <mods:namePart> subelement with type set to “date”. Other valid subelements or attributes within the <mods:name> element may be used.

Example:

```
<mods:name type="personal">
    <mods:namePart type="given">Ralitsa B.</mods:namePart>
    <mods:namePart type="family">Akins</mods:namePart>
    <mods:namePart type="date">1967-</mods:namePart>
    <mods:role>
        <mods:roleTerm authority="marcrelator" type="text">Author</mods:roleTerm>
    </mods:role>
</mods:name>
```

Name of Thesis Advisor

Mandatory practice: Encode information about the thesis advisor in the <mods:name> wrapper element with the type attribute set to “personal.” Encode the MARC relator term “Thesis advisor” in the <mods:roleTerm> subelement under the <mods:role> subelement. Encode the various parts of the name in the <mods:namePart> subelement. Include the type attribute in each <mods:namePart> subelement. The “given” and “family” name types are mandatory. The element <mods:name> is repeatable for thesis advisors.

Optional practice: Encode the birthdate in a <mods:namePart> subelement with type set to “date”. Other valid subelements or attributes within the <mods:name> element may be used.

Example:

```
<mods:name type="personal">
    <mods:namePart type="given">Bryan R.</mods:namePart>
    <mods:namePart type="family">Cole</mods:namePart>
    <mods:role>
        <mods:roleTerm authority="marcrelator" type="text">Thesis advisor</mods:roleTerm>
    </mods:role>
</mods:name>
```

Name of Committee Member

Optional practice: Encode information about committee members in the <mods:name> wrapper element with the type attribute set to “personal.” Encode the term “Committee member” in the <mods:roleTerm> subelement under the <mods:role> subelement. Encode the various parts of the name in the <mods:namePart> subelement. Include the type attribute in each <mods:namePart> subelement. The “given” and “family” name types are mandatory. The element <mods:name> is repeatable for committee members. Encode the birthdate in a <mods:namePart> subelement with type set to “date”. Other valid subelements or attributes within the <mods:name> element may be used.

Example:

```
<mods:name type="personal">
    <mods:namePart type="given">Jane R.</mods:namePart>
    <mods:namePart type="family">Smith</mods:namePart>
    <mods:role>
        <mods:roleTerm type="text">Committee member</mods:roleTerm>
    </mods:role>
</mods:name>
```

Name of Degree Grantor

Mandatory practice: Encode information about the degree grantor in the <mods:name> wrapper element with the type attribute set to “corporate.” Encode the name of the degree granting institution in a <mods:namePart> subelement. Use the form of the name authorized by the Library of Congress Name Authority File. Encode the name of the department that granted the degree in a <mods:namePart> subelement. Encode the MARC relator term “Degree grantor” in the <mods:roleTerm> subelement under the <mods:role> subelement.

Optional practice: Other valid subelements or attributes within the <mods:name> element may be used.

Example:

```
<mods:name type="corporate" authority="lcnaf">
    <mods:namePart>Texas A &amp;amp; M University</mods:namePart>
    <mods:namePart>Philosophy</mods:namePart>
    <mods:role>
        <mods:roleTerm authority="marcrelator" type="text">
                Degree grantor
        </mods:roleTerm>
    </mods:role>
</mods:name>
```

Type of Resource

Mandatory practice: Encode the type of resource in the <mods:typeOfResource> element.

Example:

<mods:typeOfResource>
    text
</mods:typeOfResource>

Genre

Mandatory practice: Encode the MARC genre term “theses” in the <mods:genre> element. Set the authority attribute to “marcgt.”

Optional practice: Other valid attributes within the <mods:genre> element may be used.

Example:

```
<mods:genre authority="marcgt">
    theses
</mods:genre>
```

Origin Information

Mandatory practice: Encode relevant dates for the ETD in the <mods:originInfo> wrapper element. The creation date is defined as the date the student graduates or the date the degree is conferred. The publication date is defined as the date the ETD is released to the public.

Encode the month and year of the creation date, according to ISO 8601, in the <mods:dateCreated> subelement. Set the encoding attribute to “iso8601.”

Encode the month and year of the publication date, according to ISO 8601, in the <mods:dateIssued> subelement. Set the encoding attribute to “iso8601.”

Optional practice: The day of the month may be included date encodings. Other valid subelements or attributes within the <mods:originInfo> element may be used.

Example:

```
<mods:originInfo>
    <mods:dateCreated encoding="iso8601">200408</mods:dateCreated>
    <mods:dateIssued encoding="iso8601">200412</mods:dateIssued>
</mods:originInfo>
```

Language

Mandatory practice: Encode information about the language of the ETD in the <mods:language> wrapper element. Encode the language, according to ISO 639-2b, in the <mods:languageTerm> subelement. Set the type attribute to “code” and the authority attribute to “iso639-2b.” The <mods:languageTerm> subelement is repeatable.

Optional practice: Other valid subelements or attributes within the <mods:language> element may be used.

Example:

```
<mods:language>
    <mods:languageTerm type="code" authority="iso639-2b">
        eng
    </mods:languageTerm>
    <mods:languageTerm type="code" authority="iso639-2b">
        spa
    </mods:languageTerm>
</mods:language>
```

Physical Description

Mandatory practice: Encode the physical description in the <mods:physicalDescription> wrapper element. Encode the MARC format term “electronic” in the <mods:form> subelement. Set the authority attribute to “marcform.” Encode the MIME type in the <mods:internetMediaType> subelement. Encode the digital origin (“born digital” or “reformatted digital”) in the <mods:digitalOrigin> subelement.

Optional practice: Other valid attributes within the <mods:physicalDescription> element may be used.

Example:

```
<mods:physicalDescription>
    <mods:form authority="marcform">electronic</mods:form>
    <mods:internetMediaType>application/pdf</mods:internetMediaType>
    <mods:digitalOrigin>born digital</mods:digitalOrigin>
</mods:physicalDescription>
```

Abstract

Mandatory practice: Encode the abstract in the <mods:abstract> element. Include the language attribute encoded in ISO 639-2b.

Optional practice: Valid attributes within the <mods:abstract> element may be used.

Example:

<mods:abstract lang="eng">This dissertation study presents a conceptual framework for implementing and assessing patient safety systems in healthcare institutions. The conceptual framework consists of critical processes and performance measures identified in the context of the 2003 Malcolm Baldrige National Quality Award (MBNQA) Health Care Criteria for Performance Excellence…</mods:abstract>

Subject

Mandatory practice: Encode topical subject terms in the <mods:subject> wrapper element. Encode individual terms or phrases in the <mods:topic> subelement. The <mods:topic> subelement is repeatable.

Optional practice: Controlled subject headings may be included by using the authority attribute of the <mods:topic> subelement. Other valid subelements or attributes within the <mods:subject> element may be used.

Example:

```
<mods:subject>
    <mods:topic>medicine</mods:topic>
    <mods:topic>patient safety</mods:topic>
    <mods:topic>processes and measures</mods:topic>
    <mods:topic>Baldridge framework</mods:topic>
    <mods:topic authority="lcsh">Universities and colleges</mods:topic>
</mods:subject>
```

Identifier

Mandatory practice: Encode the unique identifier in the <mods:identifier> element. The <mods:identifier> element is repeatable.

Optional practice: The type attribute may be used in the <mods:identifier> element. Other valid attributes within the <mods:identifier> element may be used.

Example:

<mods:identifier type="hdl">
http://handle.tamu.edu/1969.1/1042
</mods:identifier>

Location

Mandatory practice: Encode the location in the <mods:location> wrapper element. Encode the uniform resource locator (URL) in the <mods:url> subelement.

Optional practice: Other valid attributes within the <mods:identifier> element may be used.

Example:

<mods:location>
    http://handle.tamu.edu/1969.1/1042
</mods:location>

Degree information

Note: The MODS standard does not have elements specifically for theses and dissertations. In order to encode degree information in MODS, the <mods:extension> element is used to reference the ETD-MS XML schema.

Mandatory practice: Encode information about the conferred degree in the <etd:degree> wrapper element. Encode the degree name in the <etd:name> subelement. Use the fully spelled out form of the degree name. Encode the degree level, from the TDL vocabulary, in the <etd:level> subelement. Encode the degree discipline, from the TDL vocabulary, in the <etd:discipline> subelement.

Example:

```
<etd:degree>
    <etd:name>Doctor of Philosophy</etd:name>
    <etd:level>Doctoral</etd:level>
    <etd:discipline>Educational Administration</etd:discipline>
</etd:degree>
```

Record Information

Mandatory practice: Encode information about the MODS record in the <mods:recordInfo> wrapper element. Encode the name of the agency that created the MODS record in the <mods:recordContentSouce> subelement, with the authority attribute set to “marcorg.” Encode the month, year, and day of the creation date of the record, according to ISO 8601, in the <mods:recordCreationDate> subelement. Set the encoding attribute to “iso8601.” Encode the month, year, and day of the change date, according to ISO 8601, in the <mods:recordChangeDate> subelement. Set the encoding attribute to “iso8601.” Encode the unique record identifier in the <mods:recordIdentifier> subelement.

Optional practice: Other valid attributes within the <mods:recordInfo> element may be used.

Example:

```
<mods:recordInfo>
    <mods:recordContentSource authority="marcorg">
        TxCM
    </mods:recordContentSource>
    <mods:recordCreationDate encoding="iso8601">
        20050826
    </mods:recordCreationDate>
    <mods:recordChangeDate encoding="iso8601">
        20050826
    </mods:recordChangeDate>
    <mods:recordIdentifier>12345678</mods:recordIdentifier>
```

Outstanding issues

- Using ETD extensions in MODS
  - degree level, degree name, degree discipline
- How to code the department
- Vocabulary for degree discipline
- Record information
  - Content source
  - Identifier
- Describing “compound” ETDs with multiple files

Implementation

- Initially, convert from MARC or DC
  - Current metadata might not have all elements, or be formatted correctly, so collection metadata won't be perfect initially
- Eventually, generate MODS from submission process, auto-ingest into TDL
  - This assumes that we will develop a submission system that organically produces METS/MODS and feeds directly into TDL compatible repositories

Future projects

- METS and MODS for compound ETDs
- Rights Management Metadata for ETDs

Questions and Discussion