

MODS, METS, and other metadata standards

Sally McCallum
Library of Congress

Content

❖ MODS

- Description
- Sample applications

❖ Other metadata for digital environment

- METS

❖ How it fits together

Electronic resource environment

- ❖ **Digital became a standard form of material “overnight”**
 - **Easy to produce**
 - **Advantages over traditional forms**
 - **Volume enormous**
 - **“adding to the collection” not well understood**
 - **Preservation requirements are still not clear**
- ❖ **Need shortcut to/from MARC?**
- ❖ **Need more than descriptive metadata?**

MARC 21 derivative

- ❖ **Need simplicity because of large numbers of resources**
- ❖ **Need to use XML to take advantage of new protocols and XML tools**
- ❖ **Need close relationship to MARC 21 important because MARC 21 used worldwide - a billion record resource**

MARC 21 derivative

❖ MODS (Metadata Object Description Standard)

- XML; user friendly language based tags
- Electronic material a special focus
- Simpler and less detail than MARC 21, richer than Dublin Core
- Primary relationship is to MARC 21, but also enables deriving data from Dublin Core, ONIX, and digital objects themselves
- Does not assume use of any specific rules for description
- Shares definitions with MARC; element descriptions are reused throughout the schema

MARC 21 derivative

- **Coordinated with emerging data models – METS, FRBR**
- **Open development**
- **Rich recursion and linking**
- **Use of XML schema allows for flexibility and availability of freely available tools**

MARC – MARCXML - MODS

❖ MARC

- [245] 10\$aHelsinki :\$ba cultural and literary history /\$cNeil Kent

❖ MARCXML

```
<datafield tag="245" ind1="1" ind2="0">  
  <subfield code="a">Helsinki</subfield>  
  <subfield code="b">a cultural and literary history</subfield>  
  <subfield code="c">Neil Kent</subfield>  
</datafield>
```

❖ MODS

```
<titleInfo><title>Helsinki</title>  
  <subTitle>a cultural and literary history</subTitle>  
</titleInfo>  
<note type="statement of responsibility">Neil Kent</note>
```

Features

❖ Features of MODS

- One repeatable element for names, with main/added distinguished by a role element
- Related items may be briefly or fully described by same tags as are used for the item being cataloged
- Recursion of related item element enables clear coding of multiple levels
- Extension element enables bringing in or pointing to another schema for , e.g., copyright information.
- Emphasis is on access points over description

Authority data - MADS

- ❖ **MADS (Metadata Authority Description Schema)**
 - **Companion to MODS**
 - **XML schema for an authority element set that may be used to provide metadata about agents (people, organizations), events, and terms (topics, geographics, genres, etc.).**
 - **MADS has a relationship to the MARC 21 Authority format, as MODS has to MARC 21 Bibliographic -- both carry selected data from MARC 21**
 - **Still experimental**

Cataloging web collections at LC

- ❖ **Collection level MARC 21 record in the Online catalog**
- ❖ **Individual sites cataloged using MODS and searched on a web site**
 - **MODS data:**
 - **Derived from web site and reviewed by cataloger (e.g., title, description)**
 - **Inserted in all records (e.g., permissions, record numbers)**
 - **Cataloger supplied (e.g., subject, language)**
 - **Possible technician input?**
 - **Possible transformation of MODS record to MARC 21 in the future and load to OPAC**

LC Web archive cataloging

MARC 21 record for collection in OPAC

001 2007700187

050 00 \$a DT157.672

245 00 \$a Crisis in Darfur, Sudan, Web archive, 2006 \$h [electronic resource]

260 ## \$a Washington, DC :\$b Library of Congress \$c 2007-

520 ## \$a Selective collection of 216 Web sites, archived from March 20, 2006 to Nov. 20, 2006, relating to the humanitarian crisis in Darfur, Sudan. The archived sites are international in scope, and include those related to organizations involved in human rights, refugees, disaster relief, ...

650 #0 \$a Disaster relief \$z Sudan \$z Darfur

651 #0 \$a Sudan \$x Economic conditions \$y 1983-

710 2# \$a Library of Congress

856 40 #u <http://hdl.loc.gov/loc.natlib/collnatlib.00000011>

LC Web archive cataloging

MODS record for 1 web site in collection

```
<titleInfo><title>IntrerAction</title></titleInfo>
<genre>Web site</genre>
<originInfo><dateCaptured point="start" encoding="iso8601">20060302
  </dateCaptured>
  <dateCaptured point="end" encoding="iso8601">20061128</dateCaptured>
</originInfo>
<language><languageTerm authority="iso639-2b" type="code">eng
  </languageTerm></language>
<subject authority="lcsht"><geographic>Sudan</geographic><topic>History</topic>
  <temporal>Darfur Conflict, 2003- </temporal></subject>
<subject authority="lcsht"><topic>Disaster relief</topic></subject>
<relatedItem type="host">
  <titleInfo><title>Crisis in Darfur, Sudan, Web Archive, 2006</title></titleInfo>
  <location><url>http://hdl.loc.gov/loc.natlib/collnatlib.00000011</url></location>
</relatedItem>
<location><url usage="primary display">http://hdl.loc.gov/loc.natlib/mrva0011.0114
  </url></location>
```

Digitization projects

- ❖ **University of Chicago - MODS as data “hub”**
- ❖ **Resources to be digitized**
 - **May have MARC 21 records in OPAC**
 - **May have other formats and fullness of records**
- ❖ **Want to preserve granularity where possible for new faceted searching**
- ❖ **Easy to map different formats into MODS**

Aquifer initiative

- ❖ **Digital Library Federation project to build a metadata resource of distributed electronic material**
- ❖ **OAI protocol used for file building**
- ❖ **MODS selected for the metadata format**
 - **Institutions with Dublin Core metadata could enhance to MODS**
 - **Institutions with MARC 21 data could send data via MODS with little loss**
 - **Aquifer MODS Guidelines available from MODS web site - <http://www.loc.gov/mods>**

Sample MODS and MADS applications

University College Dublin

❖ Virtual Research Library and Archive (IVRLA) Project

- Cataloging digital material using FileMaker Pro database
- Convert to MODS using XSLT
- Store in Fedora archive system
- Using MADS to make accompanying authority records

MusicAustralia

- ❖ **Music record exchange retaining rich data**
 - **MODS as exchange format between National Library of Australia and ScreenSound Australia (who use a different metadata format)**
 - **Use of MODS allows for consistency with MARC data**
 - **<http://www.musicaustralia.org/>**

*Other metadata needed for
electronic resources*

Broader metadata

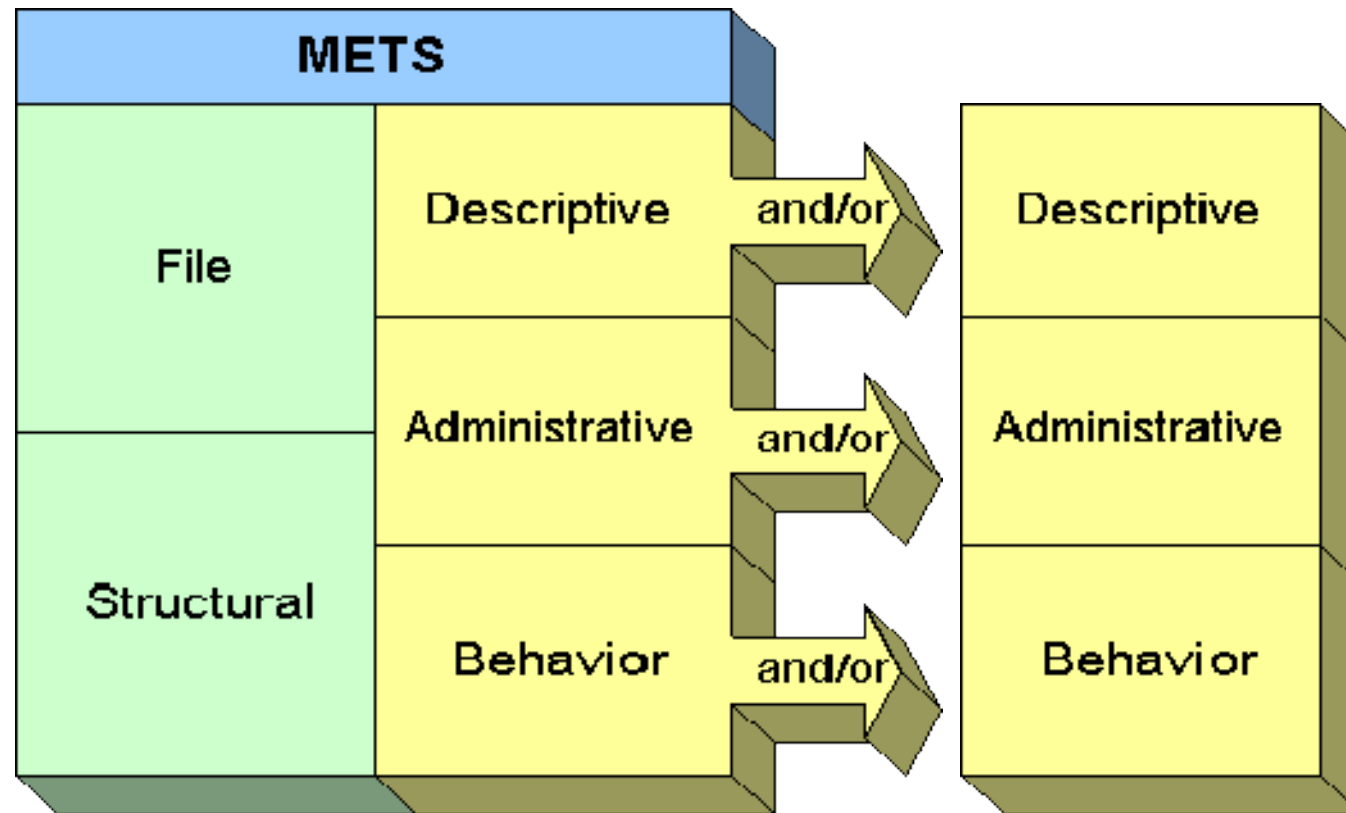
- ❖ **Descriptive metadata in MARCXML or MODS**
- ❖ **Electronic resources need more than descriptive metadata**
 - **Technical metadata (technical and structural information)**
 - **Administrative metadata (information for managing the item)**
 - **Preservation metadata (information for long-term preservation)**
 - **Rights metadata (for terms and conditions of use)**

Emerging standard - METS

❖ METS – Metadata Encoding and Transmission Standard

- XML wrapper for *descriptive AND* technical, rights, preservation, etc. metadata
- Enables resource retrieval, object validation, preservation actions, rights management, ...
- Use to submit a digital item to a repository or for interchange of digital objects
- Non-proprietary, developed by the library community
- (relatively) simple; extensible; modular
- Still need for **component standards** and profiles of usage

METS architecture



Component standard - PREMIS

❖ PREMIS – Preservation Metadata Implementation Strategies

- Evolved from projects of the 1990s
- Data dictionary of elements for core preservation metadata
- XML schema also published
- Work underway to establish best practices for using with METS
- Provides *core* preservation metadata – still need **media specific** standards

Media specific standard - MIX

❖ MIX – Metadata for Images in XML

- Technical elements needed to manage digitized image data
- Used to express attributes of digital images such as
 - file format,
 - file size,
 - dimensions,
 - resolution,
 - compression, etc.
- Recent version (1.0) includes support for GIS and JPEG 2000
- Element names harmonized with PREMIS

Media specific standard – textMD

❖ textMD – technical metadata for text

- XML Schema that details technical metadata for text-based digital objects
- allows for detailing properties such as:
 - encoding information (quality, platform, software, agent)
 - character information (character set and size, byte order and size, line terminators)
 - Languages and fonts
 - markup information
 - processing and textual notes
 - technical requirements for printing and viewing
 - page ordering and sequencing
- How do these **fit together**?

METS resource and metadata bundle

Resource Space	METS Information
Audio File (WAV, etc.)	Description MODS MARXML
Video File (MPEG, etc.)	Technical MIX, textMD
Text File (TEI, etc.)	Rights (METSRights, etc.)
Image File (TIFF, JPEG2000, etc.)	Preservation PREMIS
Web File WARC	Structure Map
	File Section

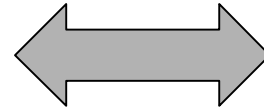
And also -

METS
MIX
PREMIS
MODS
Etc.



Repository
built from
METS
bundles

SRU
protocol



End users
and
machines
as users

Questions?

❖ Web sites for these standards:

- www.loc.gov/METS
- www.loc.gov/mix
- www.loc.gov/premis
- www.loc.gov/sru