What is XML?
- EXtensible Markup Language. XML is a set of rules for defining markup languages and describing data.

Why XML?
- XML is a standard means of delivering structured data via Web applications.
- XML is extensible—both a blessing and a burden
  - Authors can define their own tags and attributes, e.g. CML Chemical markup language
  - You may hear someone from your IT department mention "well-formed" XML. A well-formed XML file conforms to a set of very strict rules that govern XML. If a file doesn't conform to those rules, XML stops working

XML vs. HTML
- HTML tells Web browsers how to display text, images, etc.—emphasis is on display
- Unlike HTML, XML can “take database information with it”; emphasis is on structure, relationships, and ‘meaning’

A brief detour into metadata:
Two ways to designate content
In MARC: 245 04 $a The Big heat
In XML:  
  <title>Big heat</title>
  <name>value</name>

In XML the name-value pair comprises an element
An element has these parts:
  - Start tag
  - Element content
  - End tag

  <tag>content</tag>
  <subject>Goldfinches</subject>
Element rules and features

- Elements can hold data
  `<pubPlace>Boston</pubPlace>`
- Elements can hold other elements ad infinitum
  `<sourceDesc>`
  `<biblFull>`
  `<titleStmt>`
  `<title>A letter to Orestes A. Brownson</title>`
  `<author>Hildreth, Richard, 1807-1865.</author>`
  `</titleStmt>`
  `</biblFull>`
  `</sourceDesc>`
- Elements must be “properly” nested

A quick look at other XML entities

- Attributes qualify elements
  `<note type="500">Caption title.</note>`
- Document Type Definitions (DTDs) control the structure of XML documents
  `<ELEMENT note (#PCDATA)>`
  `<ATTLIST note type CDATA #IMPLIED>`
- XML Schemas give more control than DTDs
  `<xs:element ref="note"/>`
- Extensible Stylesheet Language Transformation (XSLT) stylesheets transform one XML document into another (or into HTML)

What does XML allow us to do?

- Structure data with a flexible and extensible set of rules
- Share data in a non-proprietary format, especially among “incompatible” systems
- Reuse data, e.g., in different presentation formats for different purposes

MODS

- (Metadata Object Description Schema)
- A descriptive metadata standard
- An initiative of Network Development and MARC Standards Office at the Library of Congress
- Uses XML schema
- Originally designed for library applications, but may be used for others
- A derivative of MARC

Evolution of MARC 21

- Need to take advantage of XML
  - Establish standard MARC 21 in an XML structure
  - Take advantage of freely available XML tools
- Need simpler (but compatible) alternatives
  - Development of MODS
- Need interoperability with different schemas
  - Assemble coordinated set of tools
- Need continuity with current data
  - Provide flexible transition options

MODS

- Set of 20 bibliographic elements - a subset of the MARC 21 Format for Bibliographic Data
- Not as complete as the full MARC format, but richer than Dublin Core (for example)
- Highly interoperable with existing MARC records
- Uses language-based tags, rather than numbers like MARC 21 (245, 650, etc.)
MODS

• XML-based
  – Intended to work with/complement other metadata formats
• Can be used for conversion of existing MARC records or to create new resource description records
• Useful particularly for library applications that want to go beyond the OPAC
• Shares features of MARC and Dublin Core

MODS

• An XML descriptive metadata standard
• A derivative of MARC
  – Uses language based tags
  – Contains a subset of MARC data elements
  – Repackages elements to eliminate redundancies
• MODS does not assume the use of any specific rules for description
• Element set is particularly applicable to digital resources

MODS Example

<?xml version="1.0" encoding="UTF-8"?>
<titleInfo>
<title>Sound and fury:<br>the making of the punditocracy</title>
<subTitle>the making of the punditocracy /</subTitle>
</titleInfo>

Why MODS?

• XML (Extensible Markup Language) is the markup language for the Web
• LC is investigating XML as a new more flexible syntax for MARC element set
• There is a need for rich descriptive metadata in XML, but simpler than full MARC, especially for complex digital library objects

Why use MODS?

• Not MARC
• Easier entry point
  – No codes
  – Uses XML
• Not simple Dublin Core
• MODS is richer than Dublin Core, more compatible with library data than ONIX, more end user friendly than MARCXML, and simpler than MARC

Perhaps a useful middle ground between MARC and simple Dublin Core

MODS features

• Uses natural language tags instead of numeric tags and avoids abbreviations for easier understanding
  MODS: <titleInfo><title>Sound and fury: <title>Sound and fury: <subTitle>the making of the punditocracy </subTitle> </titleInfo>

  MARC: 245 10 Sound and fury: [b the making of the punditocracy /]
• Consolidates data from different MARC tags
  – Publishing info in MARC tags 008, 044, 260 subfields a,b,c, 033, 250, 310, 321 pulled together under same MODS tag <originInfo>
MODS features

- **Physical description** info from MARC tags 008, 256, 300 and 856 are pulled together under the MODS `<physicalDescription>` tag.
- **Genre** info from MARC tags 007, 008, and 655 are combined under the one MODS tag, `<genre>`.
- MODS unique elements with no MARC equivalent:
  - `<digitalOriginal>` with values "born-digital" and "formatted-digital"
  - `<reformattingQuality>` with values "access" and "preservation", and "replacement"
- Xlink attribute allows inclusion of a link to information
- ID attribute allows inclusion of an internal link to information

MODS features

- All MODS elements and attributes are optional
- All MODS elements and sub elements are repeatable
- Some top level elements may also serve as a sub element under another element
  - Example: `<titleInfo>` can be a top level element (to show the main title of a resource) but can also function as a sub element under another top level element such as `<relatedItem>` (to show the title of another item related to a resource).

Example of element and attribute

```
<genre authority="aat"> portraits </genre>
```

Element | Attribute
---|---
Controlled vocabulary stipulated as the "authority"

Value (i.e. Controlled term)

Example of element and sub element

```
<place><placeTerm type="text">Ithaca, N.Y. </placeTerm>
```

Element | Sub elements | Attribute
---|---|---
Sub element | Text stipulated as the "type"

Value (i.e. Place name)

Example of a MODS element with sub element

- Using MODS to show series data:
  ```
  <relatedItem type="series">
  <titleInfo>
  <title>Music for voice and instrument</title>
  </titleInfo>
  </relatedItem>
  ```

- The term type functions as an attribute with the value series
- The title is given inside `<titleInfo>` and `<titleInfo>` tags

Advantages of MODS

- The metadata element set is richer than Dublin Core
- Hierarchical structure allows for rich description of complex digital objects
- The MODs element set is more compatible with existing descriptions than Dublin Core
- MODS is a rich descriptive element set that works well with hierarchical METS objects
Features of MODS

- Uses language-based tags
- Elements generally inherit semantics of MARC
- Elements particularly applicable to digital resources
- MODS does not assume the use of any specific rules for description
- Element descriptions are reused throughout the schema
- Use of XML schema allows for flexibility and availability of freely available tools

Features of MODS

- Ignores most fixed field elements, with exception of physical format and genre codes
- No required fields
- All fields are repeatable
- No main and added entries; all authors are simply authors
- Record can have multiple titles without a single “main title”

Advantages of MODS

- MODS vs. Dublin Core simple
  - Allows richer description with substructure
  - Semantics that are a more natural fit in a library environment
- MODS vs. MARCXML
  - More lightweight than MARCXML
  - Uses language based tags and more readily understandable

Potential Uses of MODS

- as an extension schema to METS
- to represent metadata for harvesting (OAI)
- for original resource description in XML syntax
- for representing a simplified MARC record in XML
- for metadata in XML that may be packaged with an electronic resource

Use in the DLF Aquifer Initiative

“to promote effective use of distributed digital library content for teaching, learning, and research in the area of American culture and life”

- Use of MODS mandatory for participation
  - Frustration with DC in OAI environment
  - Can utilize MODS to build in specific services (asset actions)
- Metadata Working Group provided guidelines for ‘shareable’ MODS records

Differences between MODS and Dublin Core

- MODS has structure
  - Names
  - Publication information (originInfo)
  - Related Item
  - Subject
- MODS is more compatible with library data
  - Semantics
  - Conversions
  - Relationships between elements
- Includes record management (administrative metadata) information
MODS high-level elements

- Title Info
- Name
- Type of resource
- Genre
- Origin Info
- Language
- Physical description
- Abstract
- Table of contents
- Target audience

MODS: Types of elements

- Simple elements
- Wrapper elements

Simple element examples

<note>March for voice and piano.</note>

<identifier type="music plate">2663 G. Schirmer</identifier>

Wrapper element example 1

<titleInfo>
  <nonSort>The</nonSort>
  <title>service flag</title>
  <subTitle>a song</subTitle>
</titleInfo>

Wrapper element example 2

<name type="personal">
  <namePart>Herschel, William</namePart>
  <namePart type="date">1873-1939</namePart>
</name>

<role>
  <roleTerm authority="marcrelator"
    type="text">lyricist</roleTerm>
</role>

Choosing MODS for descriptive metadata

MODS is particularly useful for
- compatibility with existing bibliographic data
- embedded descriptions in relatedItem
- Rich, hierarchical descriptions that work well with METS structural map
- "out of the box" schema; can use <extension> for local elements and to bring in external elements from other schemas
MODS

- MODS was developed as a rich descriptive metadata standard that works well with other XML standards/initiatives
- Provides a descriptive metadata standard intermediate in richness between Dublin Core and MARC 21 (or other specialized descriptive metadata standard)
- Although a derivative of MARC, MODS has wide potential application as a rich descriptive metadata standard for various needs
- XML tools allow for easy transformations

What is the Digital Library Federation?

an international association of libraries and allied institutions.

Mission:
to enable new research and scholarship of its members, students, scholars, lifelong learners, and the general public by developing an international network of digital libraries.

DLF Aquifer Initiative

Purpose is to promote effective use of distributed digital library content for teaching, learning, and research in the area of American culture and life.

Bringing distributed content together and pushing it out to scholars and researchers

Awarded $816,000 from the Mellon Foundation for DLF Aquifer Development for Interoperability Across Scholarly Repositories: American Social History Online

Initial aggregation at: http://quod.lib.umich.edu/a/aquifer/

OAI-PMH

- OAI PMH allows metadata to be harvested from multiple data providers
  - Usually metadata is brought together into aggregations (like OAIster)
- OAI PMH requires use of simple Dublin Core but allows use of other metadata schemas
  - Most data providers only provide Dublin Core; most aggregators only harvest Dublin Core

Difficulties in Building Services in Aggregations

- Use of Dublin Core often problematic
  - Lack of semantic complexity
- Metadata authors often are not thinking about sharing metadata outside of their local contexts
- Metadata between different collections inconsistent
- Few user studies of aggregations
- Building browse, targeted search, facets, etc… has been very, very, very difficult for aggregators

DLF Aquifer Implementation Guidelines for Shareable MODS Records

- To help create rich, shareable metadata records that is consistent and coherent
- Meant for shared records, not local implementations
- Focused on humanities and cultural heritage materials
- Based on the DLF / NSDL Best Practices for Shareable Metadata
What is shareable metadata?
- Is quality metadata
- Promotes search interoperability
  "the ability to perform a search over diverse sets of metadata records and obtain meaningful results." (Priscilla Caplan)
- Is human understandable outside of its local context
- Is useful outside of its local context
- Is machine processable

Metadata as a view of a resource
- No monolithic one-size-fits-all metadata record
- The view might be different depending on use and audience as well as format, content, and context
  - Content standard is a view
  - Metadata standard is a view
  - Vocabulary used is a view

Format of the Guidelines
- Summary of Requirements and Recommendations
- Definition from LC’s MODS User Guidelines
- Discussion of Use
- Examples
- Use by Aggregators
- Mapping to Dublin Core
- Relationship to DLF/NSDL Best Practices

Required elements
- <titleInfo><title>
- <typeOfResource>
- <originInfo> and at least one date subelement with keydate="yes"
- <language> if language primary to resource
- <physicalDescription><digitalOrigin>
- <physicalDescription><internetMediaType>
- <subject> if applicable
- <location><url usage="primary display">
- <accessCondition type="use and reproduction">
- <recordInfo><languageOfCataloging>
- <recordInfo><languageTerm>

Recommended elements
- <name><namePart>
- <genre>
- <abstract>
- <tableOfContents> if applicable
- <targetAudience> if applicable
- <note> if applicable
- <relatedItem> if applicable
- <identifier>
- <part> if applicable

Optional elements
- <classification>
- <extension>
Levels of Adoption - Tied to user tasks

1. Minimum for participation: Allows users to cite the resource
2. Minimum for doing anything useful: Allows users to perform basic searches and filtering
3. Allows more advanced functionality: Allows users to browse and group search results
4. Adopt all required guidelines (and some recommended): Allows users to perform more precise searches
5. Completely adopt all recommendations: Allows users to effectively evaluate resources

Example – Level 3

Title: Life Mask of Stephen A. Douglas
Name (authority=unaff): Leonard Wells
Role: Sculptor
Type of Resource: three dimensional object
Genre: life masks
Genre: casts (sculpture)
Date created (w3c: keyDate=“yes”): 1858
Form authority=“gmfd”: art original InternetMediaType: image/jpeg
Subject: Douglas, Stephen Arnold 1813-1861
URL usage=“primary display” access=“object in context” http://images.library.uiuc.edu:8081/tdc,252 AccessCondition type=“useAndReproduction”: Use of this public-domain resource is unrestricted.