



# Metadata standards at the Library of Congress

---

Rebecca Guenther  
AMPAS Digital Motion Picture  
Metadata Symposium  
June 11, 2009



# Overview of presentation

---

- Functions of metadata
- Types of metadata
  - Descriptive
  - Administrative (includes technical)
  - Preservation
  - Structural
- Metadata standards maintained at LC
- Implementing metadata



# Functions of metadata

---

- Discovery
- Management
- Control IP rights
- Identification
- Certify authenticity
- Mark content structure
- Indicate status
- Describe processes



# Metadata standards for digital materials

- Digital objects are complex and comprised of multiple files, e.g. scanned digital books  
<http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200033350/contactsheet.html>
- Using established standards facilitates exchange
- Complex objects require more metadata than analog for their management and use
  - Descriptive
  - Administrative
    - Technical
    - Digital provenance/events
    - Rights/Terms and conditions
  - Structural
- There is need to integrate metadata for physical and digital objects



# Types of metadata

---

- Descriptive
- Administrative
  - Technical
  - Digital provenance
  - Rights/Access
- Preservation (supported by the above)
- Structural
- Meta-metadata



# Descriptive Metadata

- Title, author, human-readable description of a resource
- Subject or topical information
- Genre and format of the resource
- Relationships with other resources (version, parent/child, etc.)



## More about descriptive metadata

---

- Most standardized and well understood type of metadata
- Increased number of descriptive metadata standards for different needs and communities
- Important for resource discovery
- May support various user tasks
- Has different aspects: content rules, controlled values, formats/schemas, syntax
- May be at various levels of granularity (e.g. the book, the article, the page, the physical item)



# The current descriptive metadata environment

- Multiplicity of descriptive metadata standards
- Cultural heritage institutions have a long tradition of exchanging metadata in a standardized form
  - MARC 21 provides an exchange standard
  - Years of record exchange has resulted in cost savings
  - Using a well-defined and understood standard allows for future migration to another system
  - Supporting standards can be leveraged in the broader metadata community
  - Many moving images have existing descriptive metadata in library catalogs and systems that can be reused





# Descriptive metadata standards maintained at LC

- Descriptive formats
  - MARC 21
  - MODS
  - EAD
- Controlled vocabulary standards at LC
  - Library of Congress Subject Headings
  - Thesaurus of Graphic Materials
  - ISO 639 Language codes
  - MARC Country and Geographic Area codes
  - MARC Relator codes
  - Library of Congress Classification



# What is MARC 21?

- A syntax defined by an international standard and was developed in the late 60s
- As a syntax it has 2 expressions:
  - Classic MARC (MARC 2709)
  - MARCXML
- A data element set defined by content designation and semantics
- Institutions do not store "MARC 21", as it is a communications format
- Many data elements are defined by external content rules
- Billions of bibliographic records world-wide



# What is MODS?

- Metadata Object Description Schema
- An XML descriptive metadata standard
- A derivative of MARC
  - Uses language based tags
  - Contains a subset of MARC data elements
  - Rich, but not as rich as full MARC
  - More compatible with existing library data than DC
- Does not assume the use of any specific rules for description
- Element set is particularly applicable to digital resources



# Administrative Metadata

- Metadata to manage the object
  - Technical metadata: technical characteristics about the object
  - Digital provenance metadata: actions that have been performed on the object
  - Rights metadata: information about access and use of the object
  - Often at a lower level of granularity than descriptive metadata



# Rights/Access Metadata

---

- Where is the resource? Is it in a place open to me?
- Are there restrictions on the use of the resource?
- What can I do with this resource?
- May include machine-actionable data used by DRM systems



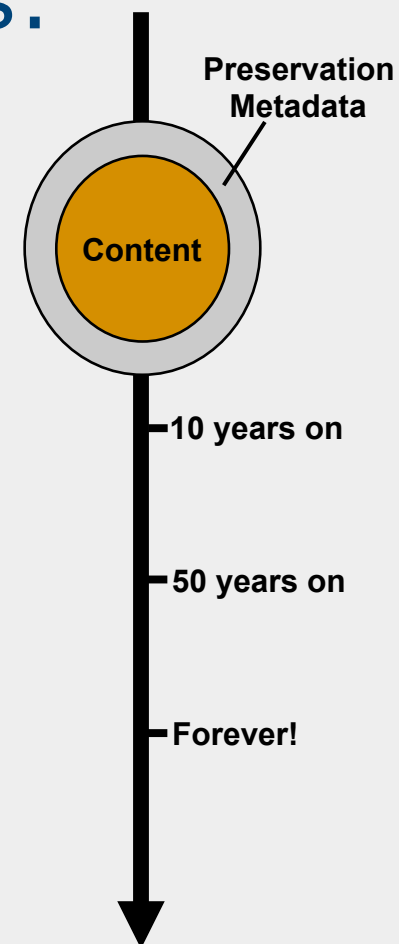
# Preservation Metadata

---

- Designed to ensure that the information the resource contains remains accessible to users over a long period of time
- Oriented toward finished products, but also applicable to works in process
- Records details about format migration and data refreshment
- Allows a variety of approaches to the problem of maintaining resources over time
- Falls into administrative metadata
- Much is extracted from object and could be carried along during the production process or machine generated

# Preservation metadata includes:

- Provenance:
  - *Who has had custody/ownership of the digital object?*
- Authenticity:
  - *Is the digital object what it purports to be?*
- Preservation Activity:
  - *What has been done to preserve it?*
- Technical Environment:
  - *What is needed to render and use it?*
- Rights Management:
  - *What IPR must be observed?*



➤ **Makes digital objects self-documenting across time**



# PREMIS Data Dictionary for Preservation Metadata

- A data dictionary for metadata to support the long-term preservation of digital objects
  - A piece of the necessary infrastructure for implementing reliable, sustainable preservation programs
- A supporting XML schema for implementation in a variety of contexts
- A maintenance activity hosted at LC including an Implementers' Group and Editorial Committee





# Technical metadata in PREMIS

---

- Object ID
- Preservation level
- Object characteristics (format, size, creating application etc.)
- Storage
- Environment (hardware and software)
- Digital signatures
- Relationships
- Linking identifiers



# Preservation actions

---

- PREMIS event information
  - Event ID
  - Event type
  - Event date/time
  - Event outcomes
  - Linking identifiers
- Need to document actions on objects for long term preservation regardless of preservation strategy



# Technical metadata for images

- Metadata for Images in XML (MIX)
  - An XML Schema designed for expressing technical metadata for digital still images
- Based on the NISO Z39.87 Data Dictionary – Technical Metadata for Digital Still Images
- Format specific metadata for images, e.g. bit depth, color space, camera settings, etc
- Most well developed of format specific technical metadata standards
- LC is maintenance agency



# Technical metadata for audio and video

- Not as well developed as other technical metadata
- Complexities of file formats requires expertise to develop these
- LC developed XML technical metadata schemas in 2003/2004 for LC Audiovisual Prototype Project; widely implemented because of the lack of other schemas
- Audio and video technical metadata schemas under development by expert organizations
- [Moving Image Collections](#) (MIC) project is also experimenting with these



# Audio/video object and provenance metadata

- Audio object metadata: AES X098B (Audio Object Schema) and AES X098C (Process History Schema)
- Many definitions could come from SMPTE-RP-210 registry of terms
- Digital provenance metadata, similar to PREMIS event
- MIC is adapting AES X098B for video



# Structural Metadata

---

- Ties the components of a complex or compound resource together and makes the whole usable
- Enables flexible and local approaches to presentation and navigation
- Various approaches to sharing structural metadata exist
- A number of container standards have been developed by different communities



# Metadata Encoding & Transmission Standard (METS)

- Developed by the Digital Library Federation, maintained by the Library of Congress
- Records the (possibly hierarchical) structure of digital objects, the names and locations of the files that comprise those objects, and the associated metadata
- To package metadata with digital object (or link to digital object) in XML syntax
- For retrieving, storing, preserving, serving resource
- For **interchange** of digital objects with metadata
- As information package in a digital repository (may be a unit of storage or a transmission format)

<http://www.loc.gov/standards/mets/>



# The structure of a METS file

**METS**

**fileSec**

file inventory

**dmdSec**

descriptive metadata

**amdSec**

administrative metadata

**behaviorSec**

behaviour metadata

**structMap**

structural map





# Other container formats

---

- Support transmission and archiving
- Bundle together metadata and content
- Some other examples:
  - Material eXchange Format (MXF)
  - MPEG 21 DIDL
  - AES 31-3 Standard for Network and File Transport of Audio - Audio-File Transfer and Exchange



# Meta-metadata

---

- Metadata about the metadata
  - Who created this information?
  - When was it created?
  - When were links last checked?
  - Other update transactions?
- May be a component of some metadata schemes
- Allows for managing the metadata, not just the resource described



# Implementing metadata

- What metadata can be carried along with the object over its lifecycle that can be used post-production?
  - Descriptive
  - Technical
  - Process history
  - Rights
  
- Metadata extraction
  - Much technical metadata may be extracted from file headers
  - Some metadata may be generated on a batch of objects

## Example of metadata for LC video object

[Great conversations: the conductors: Zubin Mehta / Eugene Istomin \[video recording\]](#)

**Title**

Great conversations: the conductors:  
Zubin Mehta [videorecording]

**Interviewee(s)**

Mehta, Zubin

**Producer(s)**

Rosen, Peter

**Editor(s)**

Warshaw, Hilan

**Host**

Istomin, Eugene

**Director**

Rosen, Peter

**Place of Publication/Creation**

New York

**Publisher(s)**

Peter Rosen Productions, Inc.

**Date issued**

2005

**Form**

videorecording



<http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200031108/default.html>

# Descriptive metadata

|                                   |   |
|-----------------------------------|---|
| <b>Title</b>                      | Great conversations: the conductors: Zubin Mehta  |
| <b>Interviewee(s)</b>             | Mehta, Zubin  |
| <b>Producer(s)</b>                | Rosen, Peter  |
| <b>Editor(s)</b>                  | Warshaw, Hilan  |
| <b>Host</b>                       | Istomin, Eugene   |
| <b>Director</b>                   | Rosen, Peter  |
| <b>Place of Creation</b>          | New York  |
| <b>Publisher(s)</b>               | Peter Rosen Productions, Inc.   |
| <b>Date issued</b>                | 2005  |
| <b>Form</b>                       | videorecording  |
| <b>Physical Description</b>       | 1 digibeta videotape; duration: 65 min., 10 sec.  |
| <b>Permissions note</b>           | Copyright Library of Congress. This program was made possible through the courtesy of Eugene Istomin. Zubin Mehta's appearance is courtesy of himself.  |
| <b>Type of Material</b>           | moving image  |
| <b>Contents note</b>              | In a series of three one-on-one discussions, the Russian-born Mstislav Rostropovich, Indian-born Zubin Mehta, and James Conlon, a native of New York City, converse about leadership and inspiration on the podium, their views about the influence of European classical music on American music, and the influence of American popular music on other cultures. |
| <b>Creator note(s)</b>            | In opening credits: "Library of Congress, Washington, D.C., presents."  |
| <b>Bibliographic history note</b> | Library of Congress extended version.<br>Additional credits: for the Music Division, Library of Congress: Jon Newsom, chief; Jan Lauridsen, assist. Chief; Ruth Foss, program specialist.   |
| <b>Repository</b>                 | <a href="#">Music Division</a>  |

# Sample PREMIS metadata for one video file

## Object metadata

**Object identifier** hdl.loc.gov/greatconv/200031108/seg01/video/0001.mpg

## Object characteristics

**Message digest algorithm** md5  
**Message digest** ceb3dbc5dacd3883d0985174ef5df7db  
**Size** 310800388  
**Format name** video/mpeg  
**Format Version**  
**Creating application** VideoLAN  
**Creating application** ffmpeg

**Relationship (structural)** (identifiers of video objects for other segments)

## Event metadata

**Event identifier** E001.1  
**Event type** validation  
**Event date/time** 2006-06-06T00:00:00.005  
**Event outcome** successful; well-formed and valid  
**Linking agent identifier** (identifier of software program that validated)

# Technical metadata for video

```
<VIDEOMD xmlns="http://www.loc.gov/VMD/" xmlns:xsi="http://www.w3.org/2001/
XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/VMD/ http://
lcweb2.loc.gov/mets/Schemas/VMD.xsd">
  <file_data>
    <bits_per_sample>24</bits_per_sample>
    <byte_order>1</byte_order>
    <compression>
      <codec_creator_app>Microsoft</codec_creator_app>
      <codec_creator_app_version></codec_creator_app_version>
      <codec_name>MPEG-4 (fast motion) </codec_name>
      <codec_quality>lossy </codec_quality>
    </compression>
    <sound>Yes</sound>
    <use>Service</use>
  </file_data>
  <video_info>
    <aspect_ratio>4x3</aspect_ratio>
    <dimensions HEIGHT="704" WIDTH="480"/>
    <duration>8 min 37 sec 647 ms</duration>
    <frame>
      <pixels_horizontal>704</pixels_horizontal>
      <pixels_vertical>480</pixels_vertical>
      <frame_rate>30</frame_rate>
    </frame>
    <note>is seekable</note>
  </video_info>
</VIDEOMD>
```

# Technical metadata for audio

```
<AUDIOMD xmlns="http://www.loc.gov/AMD/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.loc.gov/AMD/ http://www.loc.gov/rr/mopic/avprot/
  AMD_020409.xsd">
  <file_data>
    <bits_per_sample>8</bits_per_sample>
    <byte_order>1</byte_order>
    <compression>
      <codec_name>MPEG 2</codec_name>
      <codec_quality>lossy</codec_quality>
    </compression>
    <data_rate> 112 kb/s</data_rate>
    <format_name> MPEG 2</format_name>
    <sampling_frequency>44100</sampling_frequency>
  </file_data>
  <audio_info>
    <duration>8 min 37 sec 647 ms</duration>
    <note>Is seekable</note>
    <num_channels>1</num_channels>
  </audio_info>
</AUDIOMD>
```





# Closing thoughts

---

- Different kinds of metadata are needed for use/presentation and preservation of digital moving image resources
- Whatever can be carried along with the resource in various stages of its lifecycle should be saved for the future
- Collaboration between those responsible for producing and those caring for and making available these resources benefits everyone
- Preservation is a shared problem that requires a shared solution
- Standards development needs to be cooperative