Caring for Digital Materials Webinar 3: Describe it so you can find it: Metadata, finding aids, and digital asset management

Instructor: Danielle Cunniff Plumer
danielle@dcplumer.com

Series Goals

1. Participants will have a better understanding of the inherent fragility of digital objects
2. Participants will acquire information to help them select preservation formats, metadata, and backup systems for digital objects
3. Participants will be able to identify one or more actions that can be taken to improve their institution’s digital preservation efforts

Sessions

We are here:

<table>
<thead>
<tr>
<th>Session Title</th>
<th>Instructor/Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to Metadata</td>
<td>Lauren Gooday</td>
<td>Tues. April 2, 2013</td>
</tr>
<tr>
<td>Types of Metadata</td>
<td>Jacob Podell</td>
<td>Thurs., April 4, 2013</td>
</tr>
<tr>
<td>Collections Inventories and Finding Aids</td>
<td>Danielle Plumer</td>
<td>Tues., April 9, 2013</td>
</tr>
<tr>
<td>Descriptive Metadata</td>
<td>Jefferson Bailey</td>
<td>Weds., April 10, 2013</td>
</tr>
<tr>
<td>Administrative Metadata</td>
<td>Liz Bishoff and Tom Goussen</td>
<td>Mon., April 15, 2013</td>
</tr>
<tr>
<td>Preservation Metadata - PREMIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata Quality</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Session Outline

- Introduction to Metadata
- Types of Metadata
  - Collections Inventories and Finding Aids
  - Descriptive Metadata
- Administrative Metadata
  - Identifiers
  - Technical Metadata
  - Preservation Metadata - PREMIS
- Metadata Quality

Metadata

- "Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource."
- Key Concepts:
  - Structured information
  - Ease of use
  - Formal standards

Metadata Myth

- History museums don’t have "metadata," because objects don’t have titles. Metadata is for art museums and libraries.
Caring for Digital Materials Webinar 3:
Describe it so you can find it: Metadata, finding aids, and digital asset management

**Dublin Core Elements**

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>Subject</td>
<td>Language</td>
</tr>
<tr>
<td>Contributor</td>
<td>Coverage</td>
<td>Type</td>
</tr>
<tr>
<td>Publisher</td>
<td>Source</td>
<td>Format</td>
</tr>
<tr>
<td>Date</td>
<td>Relation</td>
<td>Rights</td>
</tr>
</tbody>
</table>

**PastPerfect to Dublin Core**

**Metadata Guidance**

- A good object has associated metadata.
- A good object will have **descriptive** and **administrative** metadata, and compound objects will have **structural** metadata to document the relationships between components of the object and ensure proper presentation and use of the components.

**Types of Metadata**

- Descriptive
- Structural
- Administrative
  - Technical
  - Rights Management
  - Preservation

**Functions of Metadata**

- Discover resources
- Manage documents
- Control intellectual property rights
- Identify versions
- Certify authenticity
- Indicate status
- Mark content structure
- Situate geospatially
- Describe processes
- Others?
Caring for Digital Materials Webinar 3: Describe it so you can find it: Metadata, finding aids, and digital asset management

Instructor: Danielle Cunniff Plumer

Using Metadata
- **Description**
  - To (more or less) uniquely identify an item
  - To identify the parts of an item and their relationship to the whole

Using Metadata
- **Location**
  - To show where to find the item
    - Call number
    - Archival container
    - Storage unit
    - Uniform Resource Indicator

Using Metadata
- **Condition**
  - To document the condition of an item at a given time
  - To record any actions taken with respect to the item’s condition

Using Metadata
- **Use**
  - To explain conditions of use for an item
    - Based on condition
    - Based on rights

Digital is Different
- Does the metadata describe the physical item or the digital item, or both?
  - Physical item
    - Metadata as “surrogate” for the physical item
    - Metadata for inventory of and access to the physical item
    - Metadata aggregated for use in “union catalogs”
  - Digital item
    - Metadata as a component of the digital object itself
    - Metadata used as a way of pointing to the digital object from metadata aggregations

Metadata and Digital Objects
- Metadata can be:
  - In the digital object
    - File headers, e.g., TIFF, EXIF; EAD, TEI headers, Broadcast WAV
  - Near the digital object
    - Same directory, hard drive, network, asset management system
  - Far from the digital object
    - On another network, in another state, in another country
NISO’s Metadata Principles


1. Good metadata conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection.

2. Good metadata supports interoperability.

3. Good metadata uses authority control and content standards to describe objects and collocate related objects.

4. Good metadata includes a clear statement of the conditions and terms of use for the digital object.

5. Good metadata supports the long-term curation and preservation of objects in collections.

6. Good metadata records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification.

Collections Inventories

- A list of items in your collection
- Can be a Word document, an Excel file, or a database (collections management software)
- Includes information about the items
  - Name
  - Location
  - Dates
  - Rights

Collections Inventory (Example 1)

Fields:
- Collection Title
- Collection Number
- Series
- Description
- Fuller description
- Box/Folder
- Item_level description
- Information on content
- Date
- Rights
- Physical check

Collections Inventory (Example 2)

Fields:
- Box #
- 099: Local Call Number
- Photographer
- 1xx: Creator
- 1xx: Authorized Heading?
- 245: Collection Title
- 245 $f/q: Collection Date
- Extent
- Subject
- 300: Physical Description
- 340a: Physical Medium
- Note

Questions?
Caring for Digital Materials Webinar 3: Describe it so you can find it: Metadata, finding aids, and digital asset management

Instructor: Danielle Cunniff Plumer

Finding Aids
- A guide to a single collection
- Includes information about the collection and its creator(s)
- Includes information about access to and use of the collection
- May include a list of contents
  - Organized by:
    - Series > Subseries > Box > Folder > Item

Finding Aid Example

Encoded Archival Description
- XML schema or DTD that encodes the parts of a finding aid
- Finding aid can then be converted to HTML for online display
- Software to produce:
  - XML Editor (XMetal, Oxygen)
  - Archon
  - Archivist’s Toolkit
  - Excel (see http://orbiscascade.org/index/northwest-digital-archives-tools for instructions)

Library Catalogs
- Not just for books!
- Descriptive metadata (bibliographic)
- Include holdings and location information, item status, and more
- May include links to electronic materials and thumbnail images
- Increasingly tied to systems for managing access privileges

Poll: Your existing tools
- What do you already use for some or all of your digital collections?
  - Collections Inventory
    - Excel, Word, Access or other database
  - Finding Aid
    - Word, EAD, or XML
  - Dedicated collection management software
    - Library catalog, PastPerfect, Archon/AT, etc.
  - Digital collections not currently included in collections plan

Questions?
Descriptive metadata
- Most standardized and well understood type of metadata
- Different descriptive metadata standards for different needs and communities
- The actual information contained in the descriptive metadata should be developed according to community content standards

What is a Content Standard?
- Describes the types of data to be recorded in describing an entity
  - Title
  - Creator
  - Edition or version
  - Publication
  - Identifier
  - Terms of availability
- May provide information on where to find the data and how to format it

Common Content Standards
- **AACR2** (Anglo-American Cataloging Rules)
  - Content standard for traditional library cataloging
  - Closely tied to MARC schema/syntax
- **RDA** (Resource Description and Access)
  - Successor to AACR2
  - Aspires to be independent of a particular syntax
- **DACS** (Describing Archives: a Content Standard)
  - Used in archives
  - Relies on AACR2 in many areas
- **CCO** (Cataloging Cultural Objects)
  - Newer standard developed by visual arts and cultural heritage community

Descriptive Metadata Schemas
- **Dublin Core**
  - Simple set of elements
  - Designed for digital objects
- **IPTC Core**
  - Used in digital photographs
- **MARC**
  - Used in library catalogs
  - Closely tied to AACR2 content standard
- **MODS**
  - XML schema for bibliographic-type data
- **PBCore**
  - Extends Dublin Core to add elements for audiovisual assets, including technical metadata

Dublin Core Elements

<table>
<thead>
<tr>
<th>Descriptive Metadata Elements</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Description</td>
<td>Identifier</td>
</tr>
<tr>
<td>Creator</td>
<td>Subject</td>
<td>Language</td>
</tr>
<tr>
<td>Contributor</td>
<td>Coverage</td>
<td>Type</td>
</tr>
<tr>
<td>Publisher</td>
<td>Source</td>
<td>Format</td>
</tr>
<tr>
<td>Date</td>
<td>Relation</td>
<td>Rights</td>
</tr>
</tbody>
</table>

Poll: Metadata Elements
- Which descriptive metadata elements does your institution regularly collect? Check all that apply.
  - Identifier
  - Title
  - Creator/Author
  - Date
  - Description
  - Subject
  - We don’t collect any metadata
Questions?

Structural metadata

- Supports the intended presentation and use and navigation of an object
- Binds the parts together; expresses relationships between parts of a multipart object
- Examples of structural metadata needs:
  - Books, newspapers (multi-page documents)
  - Items with multiple "views" (front, back, etc.)
- Types: PDF, METS, "compound objects"

Structural Metadata in XML

Administrative metadata

- Provides information to help manage a resource
  - Preservation metadata
    - Content, fixity, provenance, context
    - Information about actions on an object
  - Technical metadata
    - Format, extent
    - Object creation information (EXIF)
  - Rights metadata
    - Access rights and restrictions
    - Preservation rights and restrictions

Dublin Core Element Set

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
<th>Identifier</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contributor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Identifier: Best Practice

- A good object will be named with a persistent, globally unique identifier that can be resolved to the current address of the object.
  - Good identifiers will at minimum be **locally unique**, so that resources within the digital collection or repository can be unambiguously distinguished from each other.
  - Global uniqueness can then be achieved through the addition of a **globally unique prefix** element, such as a code representing the organization.

Source:
Uses of Identifiers

- The description of a digital object must be specific enough to distinguish it from similar objects
- Unique identifiers make this easier
- The link between a digital object and its metadata must exist and be maintained over the entire lifespan of the object
- Persistent identifiers are essential

Identifiers and Object Names

- Use only alpha-numeric characters for both files and folders
- Exceptions are dashes (-) and underscores (_)
- Do not use special characters
- Examples include: / > < + = ^ | { [ ] } 
- These characters are used by the operating system
- Use a three character file extension (i.e. "tif" ".pdf", etc.)
- See http://www.openwith.org/ for common file extensions and the programs used to open them
- Do not use spaces in file/folder names – use dashes or underscores instead
- Note: store full path information along with file name

Identifiers and Object Names

Checksums as Identifiers

- Checksum: an identifier computed algorithmically from the contents of a digital file
- Data Fixity Check: take a "snapshot" of the contents of a digital file
- File Integrity validation: re-calculate the checksum periodically and compare it to the original

Technical Metadata

- Information about the digital object
- Often automatically created when the object is created; embedded in object header
- Exif
  - Date/time
  - Camera
  - Camera Settings

Technical Metadata Dictionary

  - http://www.niso.org/psog/group_public/download.php/6502/
Caring for Digital Materials Webinar 3: Describe it so you can find it: Metadata, finding aids, and digital asset management

Instructor: Danielle Cunniff Plumer

Controlled Vocabularies
- Format
  - Note: Not always the same as the file extension
  - Use MIME types:
    - http://www.iana.org/assignments/media-types/
- Language
  - Use RFC 4646
- Type
  - Use Dublin Core Type vocabulary
- Date
  - Use a defined standard such as ISO 8601
    - See http://www.loc.gov/standards/datetime/
    - for a draft of the Extended Time/Date Format in development at the Library of Congress

Saving Technical Metadata
- Save metadata in digital object headers
  - Microsoft Windows: Properties > Details
  - BWF MetaEdit: Broadcast WAVE metadata editor
    - http://bwfmetaedit.sourceforge.net/
  - Image Editing Software (Adobe Bridge, Aperture)
- Extract metadata from headers
  - JHOVE
    - http://hul.harvard.edu/jhove/distribution.html
  - New Zealand Metadata Extractor
    - http://meta-extractor.sourceforge.net/
- Save metadata as text (XML) files with digital objects
  - XMP metadata “sidecars”

Embedded Metadata
- Federal Agencies Digitization Guidelines Initiative (FADGI) Recommendations
  - Minimal Descriptive Embedded Metadata in Digital Still Images
  - Embedded Metadata in Broadcast WAVE Files
    - http://www digitizationguidelines gov/guidelines/digitize-embedding.html
  - Embedded Metadata in TIFF Images
    - http://www digitizationguidelines gov/guidelines/digitize-tiff.html

Editing Metadata

Questions?

Preservation metadata
- 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 intellectual property restrictions must be observed?
Caring for Digital Materials Webinar 3: Describe it so you can find it: Metadata, finding aids, and digital asset management

**PREMIS Data Dictionary**
- 224-page report includes:
  - Background
  - PREMIS Data Dictionary 2.0

http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

**PREMIS Scope**
- Preservation metadata
  - Descriptive metadata is out of scope
- Technical metadata applying to all or almost all format types
  - Format-specific metadata is out of scope
- Rights information for preservation actions, not for access

**What PREMIS Is**
- A common data model for thinking about preservation metadata
- A set of terms with definitions to help ensure consistency across implementations
- A framework of guidance for local implementations
- A standard for exchanging information packages between repositories

**What PREMIS Is Not**
- An out-of-the-box preservation solution
  - PREMIS data must be instantiated as custom metadata elements in a specific repository
  - Each use of PREMIS should be documented in an application profile
- A complete preservation solution
  - PREMIS provides preservation metadata only
  - Does not assist in the lifecycle management of objects themselves, particularly outside of a specific repository
- A complete rights management solution
  - PREMIS rights management is limited to permissions regarding actions taken within a specific repository

**PREMIS Entities**
- **Object**
  - Object ID
  - Preservation level
  - Object characteristics (format, size, etc.)
  - Storage
  - Environment
  - Digital signatures
  - Relationships
  - Linking identifiers

**PREMIS Entities**
- **Event**
  - Event ID
  - Event type
  - Event date/time
  - Event outcomes
  - Linking identifiers

- **Agent**
  - Agent ID
  - Agent name

- **Rights**
  - Rights statement
  - Granting agent
  - Permission granted
Caring for Digital Materials Webinar 3: Describe it so you can find it: Metadata, finding aids, and digital asset management

Instructor: Danielle Cunniff Plumer

Rights metadata
- Rights information is not well understood
- Different laws in different jurisdictions
- Machine actionable vs. human understandable
- Rights take different forms
  - legal statutes, e.g. copyright, privacy
  - contractual rights, e.g. licenses
- Different institutions have different rights
  - Libraries and archives have "preservation exemption" in U.S. copyright law that museums do not have

Copyright Resources
- Copyright tools:
  - Copyright slider: http://librarycopyright.net/resources/digitalslider/
  - Section 108 (Preservation Exemption) Spinner: http://www.librarycopyright.net/resources/spinner/
  - Fair Use Evaluator: http://librarycopyright.net/resources/fairuse/

Rights schemas
- PREMIS Rights
  - Focused on rights for preservation rather than access
  - Revision of PREMIS data dictionary expanded this area
  - Allows for extensibility, i.e. inserting another rights schema
- Creative Commons
  - Allows creators to choose a license for their work
  - Simple rights statements that fit a lot of situations
  - CC0 – public domain
  - http://creativecommons.org/

Poll: Preservation Metadata
- Select the preservation metadata elements that you currently use or plan to use to assist with digital preservation:
  - Format
  - Type
  - Checksums
  - Rights
  - PREMIS Events
  - Provenance
  - Other
  - N/A

Questions?

Metadata Quality
- Completeness
- Accuracy
- Provenance
- Conformance to expectations
- Logical consistency and coherence
- Timeliness (Currency and Lag)
- Accessibility

Caring for Digital Materials Webinar 3:
Describe it so you can find it: Metadata, finding aids, and digital asset management

Improving Metadata Quality
- Documentation
  - Basic standards, best practice guidelines, examples
  - Exposure and maintenance of local and community vocabularies
  - Application Profiles
  - Training materials, tools, methodologies

Metadata Quality Tools
- OpenRefine
  - http://openrefine.org/
  - Formerly Google Refine
- Free Your Metadata
  - http://freeyourmetadata.org/
    - Clean up
    - Reconcile
    - Entity Extraction

Contact Me
- Danielle Cunniff Plumer
dcplumer associates
  - http://dcplumer.com
danielle@dcplumer.com