Audiovisual and Preservation Metadata

NDSR-NY at MoMA Case Study

Peggy Griesinger

Outline

- Introduction to PREMIS
- Introduction to A/V metadata standards
- Overview of NDSR project
- Identifying metadata requirements
- Local metadata implementation
- Workflow design

PREMIS

PRESERVATION METADATA MAINTENANCE ACTIVITY

- http://www.loc.gov/standards/premis/v3/index.html
- Has four basic units:
  - Object
    - Can be Intellectual Entity, Representation, File, or Bitstream
  - Event
  - Agent
  - Rights
- You link these together to create the history of a digital object

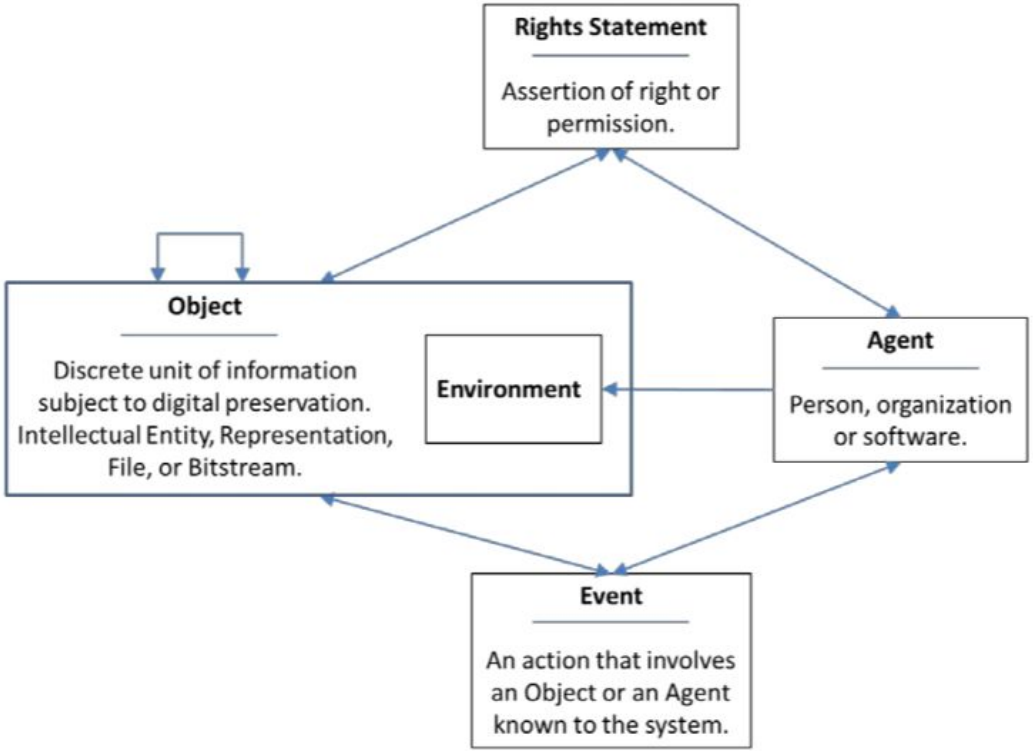

Figure 1: The PREMIS Data Model

- http://pbcore.org/elements/
- Designed to expand on Dublin Core for public broadcasting archives
- Very specifically designed for audiovisual materials
- Uses idea of “instantiations” of an intellectual object
  - Describes the intellectual content of an object
  - Can also describe technical details of:
    - Master preservation copy
    - Viewing copy
    - Original broadcast tape
    - Etc.

PBCore Example

- Intellectual entity/asset: The Shining (1980 Motion Picture)
  - Title, creator, publisher, subject, genre, etc.
- Instantiation 1: VHS copy
  - Technical information such as tape size, audio configuration, duration, etc.
- Instantiation 2: DVD copy
  - Technical information such as audio configuration, aspect ratio, language tracks, etc.

EBUCore

- https://tech.ebu.ch/MetadataEbuCore
- Also designed to expand on Dublin Core for public broadcasting archives, but created and maintained by European groups
- Has some conceptual differences from PBCore, but largely similar
- Has a very advanced RDF ontology
    - Probably better choice than PBCore if you're doing linked data work, although you can map PBCore to RDF using the EBUCore Ontology: https://pbcore.org/mappings.html

reVTMD

- https://www.weareavp.com/products/avi-metaedit-revtmd/
- Obscure standard with very specific purpose
- Addresses the need to record process history for audiovisual materials
- Describes - in granular detail - device chains used to digitize or migrate time-based media
- Generally speaking, not robust enough or well-supported to be used as a standalone

reVTMD Example

- Capture history (example: migration from VHS to .mp4 file)
  - Device 1 (called “coding process history” in reVTMD)
    - Role
    - Manufacturer
    - Serial number
    - Settings
    - Etc…
  - Device 2
  - Device 3
  - Etc…

Technical standards

- Numerous tools used in the processing of time-based media generate detailed technical output:
  - MediaInfo
  - ffprobe
  - Etc.
- These can be incorporated into records that require a human touch for other data input
- Many of them can be mapped to existing standards like PBCore and EBUCore

NDSR at the Museum of Modern Art (MoMA)

- Time-based media conservation
- Binder: customized digital repository software
  - Uses customized Access to Memory (AtoM) and Archivematica
  - Generates detailed technical metadata upon ingest
- Problem: documenting process history

https://github.com/artefactual/binder

Process History

Original ¾ inch
U-matic tape

10-bit uncompressed file
for preservation

Digital file for
exhibition use

Digital Betacam tape for
preservation

Process History

Identifying Metadata Requirements

- Existing documentation
  - What metadata already exists, in whatever format?
  - Is there a way to systematically migrate that existing data to a standardized format, or will it require more hands-on intervention?
  - Will the data require significant transformation to adhere to an existing standard, or is it easily mapped?
- Identify the gap and research standards
  - What metadata is missing?
  - What existing standards can address this gap?
    - Is one sufficient, or do you need to combine a few together?
    - If using multiple standards, how will you organize them?

Identifying Metadata Requirements

- Conversations with stakeholders
  - Focus on both:
    - Staff doing metadata creation
    - Any stakeholders who will be working with a front end version of the data
  - Development of use cases
    - Work with metadata creators to understand why they need this metadata and the ways they will potentially use it
    - Examine existing workflows to understand how stakeholders are using/creating any existing metadata
  - What existing standards can potentially provide solutions for these use cases?

Identifying Metadata Requirements

- Realistically, what resources does your institution have to sustain this metadata?
  - Will this add to the existing workload of already overloaded staff?
  - Do you or someone else have the time to take on ongoing maintenance and quality control?
  - How much experience do the metadata creators have with practical metadata creation or equivalent skills?
  - You may have a dream of a beautiful perfect metadata implementation that is not actually feasible for your institution
    - Create the best implementation you can in your circumstances
    - No matter what, avoid blatantly breaking the rules of a standard
      - Opt instead to omit information or keep it in a strictly local field

Local Implementation

- How will you interpret the broad rules of the standard?
  - Standards offer guidelines but leave room for interpretation for local needs
  - You need to make these choices *now* so your data is consistent
- What are the required fields for your purposes?
  - For example: PREMIS only requires a small handful of fields - based on your institution's resources, you may only be able to aim for filling out those, but your preservation requirements may demand more
  - It might be very important to your collection to always have the physical format recorded in detail, or you might not care about this
    - E.g., physical details of a wax cylinder recording vs. a CD-ROM

Local Implementation

- What controlled vocabularies fit your needs?
  - PBCore or EBUCore vocabularies?
  - Getty Art & Architecture Thesaurus (AAT)?
  - LCSH?
  - A local controlled list created by staff?
  - There are a LOT of vocabularies out there, so there is likely something already in existence that serves your needs

Local Implementation

- What should metadata creators do when they can't find a term in your selected vocabulary?
  - If you need to make your own terms, come up with rules for forming them
    i.e., for names:
    - Personal names – Last name, first name, middle name or initial, birth year-death year – use all when applicable
    - Corporate names – natural order
  - If you come up with your own terms, record them somewhere so you reuse them appropriately

Local Implementation

- What should metadata creators do when important information doesn't fit into the standard as implemented locally?
  - If this is a major change that will affect a large amount of records, it's likely a reconfiguration for the best practices documentation will be needed
  - If it's a small change or something that will only affect a handful of records, have a designated "catch-all" field like "description" or "notes" where miscellaneous information can be put
  - Try to anticipate these types of things while you're defining metadata requirements - look for edge cases or design an "exit strategy" to deal with unforeseen circumstances like these

Local Implementation

- Will you follow a data content standard like RDA or DACS?
  - Difficult to follow these exactly with digital collections
  - Use them as a guide to inform knowledge organization
- Should data entry be standardized for any field?
  - Absolute must for date fields
  - Best practice for dates: use ISO 8601
    - YYYY-MM-DDThh:mm:ss.sTZD
    - (eg 1997-07-16T19:20:30.45+01:00)
    - Or just 1997-07-16 for most purposes
  - Could also be useful when recording technical information about digital files
- **Clear up any ambiguity about how you're using terms**

Workflow Design

- Documentation
  - Be so detailed people think you have too much time on your hands
  - Create guidelines that eliminate as much ambiguity as possible, but don't be so strict with rules that you leave no room for the natural growth of your collection and metadata needs
    - Reassess your documentation after metadata creation has occurred for a short period of time
    - Is the documentation working in practice, or does it need to be adjusted?
    - Do this with care - constantly changing your documentation will lead to inconsistency in the project
  - If you've made unconventional choices, explain them so future staff understand you made those choices intentionally

Best Practices Documentation

- List out the following:
  - Every element
  - Every subelement
  - For more complex implementations, every attribute (for each element and subelement)
- Indicate whether each is repeatable/non-repeatable (R/NR)
- Indicate whether each is optional or required/mandatory (O/R)
- For each element, subelement, and attribute:
  - Give a basic description of what this field should be used for
  - Define the type of data to enter in it (i.e., how one should format a date)
  - Link to or list any preferred or required controlled vocabularies

Example Element

Title (non-repeatable, required)
Formal titles: For formal titles: transcribe a title as it appears on the source of information.
Supplied titles: A supplied title is one provided by the archivist when there is no formal title for the materials being described, or where the formal title is misleading or inadequate. Taking the information from any reliable source, compose a brief title that uniquely identifies the material. If using a title that repeats, number the titles so that each title is unique.

For letters, use the format: Name of Author to Name of Recipient, Full Date. For example: Peggy Griesinger to Beyoncé Knowles-Carter, December 4, 2018.

Speeches: Name of Speaker(s) at Event/Venue/Location of Speech, Date. Example: Peggy Griesinger at Corporate Annual Summit, January 12, 2019.

Workflow Design

- Quality control
  - Intermittently during the metadata creation process, have someone check the metadata for typos, errors, and/or incorrect application of the standard
  - Programs like **OpenRefine** can help with cleaning data
    - Excel can do a lot of the stuff that OpenRefine can, but I find it less intuitive for non-experts
  - Don't just fix people's errors – tell them (nicely) where they have problems so they can improve their metadata creation skills
- Metadata storage
  - This will largely depend on your digital repository/archive options at your institution
  - Best practice is that metadata should be stored alongside digital objects, preferably in a non-proprietary, standardized format like XML or JSON

Metadata Design and Implementation

1. Decide what information you need to record.
2. Select a standard or standards to use based on what you need to record.
3. Decide on and thoroughly document your local implementation of the standard.
4. Ensure that metadata creators adhere to the local implementation with quality checks.
5. Preserve and store the documentation alongside the collection.

Questions?

Peggy Griesinger

Metadata Technologies Librarian

University of Notre Dame, Hesburgh Libraries

peggygriesinger@gmail.com

Project info: https://github.com/peggygriesinger/NDSR