Preservation Planning and PREMIS

Lisa Gregory
State Library of North Carolina
April 2013
1. Planning for preservation
2. First steps towards implementation
3. PREMIS overview
Handout (URLs)

http://1.usa.gov/ZwnYKx

> Then, see top of page for handout
INDIVIDUAL WORKSTATIONS
EXTERNAL Drives
REMOVABLE MEDIA
SHARED LOCAL STORAGE
MOBILE CONTENT MANAGEMENT SYSTEMS
LAB/SHARED WORKSTATIONS
INSTITUTIONAL BACKUPS
MOBILE DEVICES
INTERNAL/EXTERNAL SERVICES
HOT
MESS
• The Signal: http://blogs.loc.gov/digitalpreservation/
• Twitter: search on [“digital preservation” or “digitalpreservation”] and start following those who post
• D-Lib Magazine
Assess
Assess

- Preservation statements: [http://www.dlib.org/dlib/january13/webb/01webb.html](http://www.dlib.org/dlib/january13/webb/01webb.html)
Organizational Infrastructure

Example: “A3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve. “

Digital Object Management

Example: “B2.5 Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs).”

Technologies, Technical Infrastructure & Security

Example: “C1.2 Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository’s services and for the data held, e.g., metadata associated with access controls, repository main content.”
If you don't write it down it never happened.
• Write down...
  – What you’re promising to stakeholders
  – Results of assessment
  – Workflows
• Be transparent
• Bring in as many involved parties as is feasible
• Yes, you’ll need to review it
Preservation Policy Template
Digital Preservation Policy & Planning Workshop

2010-10-15

Policy Statement
This is a simple statement that relates digital preservation to the institution’s mission and the communities it serves.

Summary Statement (try to keep to 1/2 page total)
This is a set of simple paragraphs that summarize the overall intent of the institution. May include the following elements:
- What are digital resources?
- Who makes selection decisions?
- When should decisions regarding preservation be made?
- What forms are prioritized (masters vs use copies)?
- Where does metadata fit?
- Why does it preserve content (e.g., institutional, legal, consortial obligations)?
- Who wrote this policy?
- How often is this policy re-evaluated and by whom?

Scope
This section summarizes the resource groups (e.g., units, departments, or external parties) for which the institution takes responsibility and prioritizes these according to institutional importance.

Selection Criteria
This section outlines the way decisions are made regarding what will be preserved. May include the following:
- Mission
- Importance to user communities
- Quality of the digital resource
- Risk to the items/collections
I WANT TO BE YOUR FRIEND

Communicate
Implement
Implement

- Storage
- Obsolete media
- Organization
- Workflows
- Metadata
Implement: Storage

- You know where stuff is, now...
  - Is what you’re taking in matching your resources?
  - Is it organized according to a system you’ve defined?
  - Who can access/write to what parts?
• You know what you have, now...
  – Do you have the staff expertise and equipment to extract/work with the data?
  – Can you at least determine if the content is viable/relevant to your collecting scope?
  – Have you articulated what you will promise about such media going forward?
Implement: Organizational Infrastructure

• You’ve written your policies, now...
  – How do you position yourself for administrative buy-in?
  – What could you reasonably improve within the next 3-5 years?
  – Are there other stakeholders headed your way?
• You’ve figured out what you currently do, now...
  – What small improvements could be made?
  – Are there tools that could help diminish the number of times a file is handled?
  – Can you start to think of an ideal workflow to work toward?
Implement: Metadata

• You’ve got your files, now...
  – What can you do to start recording/improve your preservation metadata?
  – Have you created a data dictionary or guidance document for that metadata?
  – Do you have audit trails for what happens to your files?
PREMIS

Data Model

- Defines semantic units
- Describes constraints for use
- Gives examples

Data Dictionary

- 5 entities
- Organizes semantic units
- Defines relationships

- Defines semantic units
- Describes constraints for use
- Gives examples
Alexander Calder, *Hanging Mobile*, 1951
Data Model: Example

- Inauguration speech [Intellectual entity]
- Speech2013v5.docx [object]
- Collection’s copyright & use statement [rights]
- JHOVE [agent]
- File format identified and validated [event]
# PREMIS Data Dictionary: Main Semantic Units

1. Object Identifier
2. Object Category
3. Preservation Level
4. Significant Properties
5. Object Characteristics
6. Original Name
7. Storage
8. Environment
9. Signature Information
10. Relationship
11. Linking Event Identifier
12. Linking Intellectual Entity Identifier
13. Linking Rights Statement Identifier
**PREMIS Data Dictionary:**

<table>
<thead>
<tr>
<th>Semantic unit</th>
<th>1.6 originalName</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic components</td>
<td>None</td>
</tr>
<tr>
<td>Definition</td>
<td>The name of the object as submitted to or harvested by the repository, before any renaming by the repository.</td>
</tr>
<tr>
<td>Rationale</td>
<td>The name used within the preservation repository may not be known outside of the repository. A depositor might need to request a file by its original name. Also, the repository may need to reconstruct internal links for dissemination.</td>
</tr>
<tr>
<td>Data constraint</td>
<td>None</td>
</tr>
<tr>
<td>Object category</td>
<td>Representation</td>
</tr>
<tr>
<td>Applicability</td>
<td>Applicable</td>
</tr>
<tr>
<td>Examples</td>
<td>N419.pdf</td>
</tr>
<tr>
<td>Repeatability</td>
<td>Not repeatable</td>
</tr>
<tr>
<td>Obligation</td>
<td>Optional</td>
</tr>
<tr>
<td>Creation / Maintenance notes</td>
<td>This value would always be supplied to the repository by the submitter or harvesting application. How much of the file path to preserve would be up to the repository.</td>
</tr>
<tr>
<td>Usage notes</td>
<td>This is the name of the object as designated in the Submission Information Package (SIP). The object may have other names in different contexts. When two repositories are exchanging content, it would be important for the receiving repository to know and record the name of the representation at the originating repository. In the case of representations, this may be a directory name.</td>
</tr>
</tbody>
</table>
## PREMIS Data Dictionary: Object Categories

<table>
<thead>
<tr>
<th>Object category</th>
<th>Representation</th>
<th>File</th>
<th>Bitstream</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicability</td>
<td>Applicable</td>
<td>Applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Examples</td>
<td></td>
<td>N419.pdf</td>
<td></td>
</tr>
<tr>
<td>Repeatability</td>
<td>Not repeatable</td>
<td>Not repeatable</td>
<td></td>
</tr>
<tr>
<td>Obligation</td>
<td>Optional</td>
<td>Optional</td>
<td></td>
</tr>
</tbody>
</table>
## North Carolina Digital Collections

### Preservation Metadata Implementation Guidelines

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Bit Depth</td>
</tr>
<tr>
<td>2.</td>
<td>Capture Tools</td>
</tr>
<tr>
<td>3.</td>
<td>Checksum</td>
</tr>
<tr>
<td>4.</td>
<td>Digital Characteristics (Access)</td>
</tr>
<tr>
<td>5.</td>
<td>Digital Creation Date</td>
</tr>
<tr>
<td>6.</td>
<td>Original Object Identifier</td>
</tr>
<tr>
<td>7.</td>
<td>Preservation File Name</td>
</tr>
<tr>
<td>8.</td>
<td>Preservation File Path</td>
</tr>
<tr>
<td>9.</td>
<td>Resolution</td>
</tr>
<tr>
<td>10.</td>
<td>Revision</td>
</tr>
<tr>
<td>11.</td>
<td>Security</td>
</tr>
<tr>
<td>12.</td>
<td>URL (Access)</td>
</tr>
<tr>
<td>Element</td>
<td>N/A</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Mandatory</td>
<td>Yes</td>
</tr>
<tr>
<td>Repeatable</td>
<td>Yes</td>
</tr>
<tr>
<td>Crosswalk to PREMIS</td>
<td>1.6 originalName</td>
</tr>
<tr>
<td>Controlled vocabulary</td>
<td>None</td>
</tr>
<tr>
<td>Definition</td>
<td>An identifier for the master copy generated by the creator/submitter and formatted as received.</td>
</tr>
<tr>
<td>Input guidelines</td>
<td>1. Enter filename of the digital object with extension.</td>
</tr>
<tr>
<td>Examples</td>
<td>P1080067.JPG</td>
</tr>
<tr>
<td></td>
<td>pubs_lawsofnorthcarol1817nort.pdf</td>
</tr>
</tbody>
</table>
Lisa Gregory
Digital Collections Manager
State Library of North Carolina
lisa.gregory@ncdcr.gov

Flickr users
Composingfun
Stephanie.says
Helena_perez_garcia
Jeremy Brooks
Zeptonn
Marcel Douwe Dekker
sillydog
somegeekintn

Webinar Feedback
http://www.surveymonkey.com/s/2013_DP1