11-2016


M Caruso
D Krewer
J Mullins
Liz Woolcott
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/lib_present

Part of the Library and Information Science Commons

Recommended Citation
Recommending community best practices for metadata in the Digital Preservation Network

A progress report by the DPN Preservation Metadata Standards Working Group: Moriah Caruso, Liz Woolcott, Drew Krewer, Jennifer Mullins
Overview

1. About DPN
3. Current and Future Directions
4. Work Accomplished So Far
5. Challenges Encountered
6. Invitation for Feedback
About DPN
About DPN

- “Dark archive”
- “Redundant and varied technical and legal infrastructure to assure the survival, ownership and management of preserved digital content”
- 50+ charter members, including public and private universities and colleges, consortia, and digital curation service providers
- Collaborative community
- Members are just starting to deposit content
Preservation Metadata Standards Working Group

- Was created to outline the metadata standards that DPN will follow for preservation.
  - Deliverable 1: Define a set of fields that will remain constant and available with all deposits made into DPN.
  - Deliverable 2: Publish a paper documenting process and relevant work done at other institutions leading to the selection of these standards.
- Started its work in January 2016.
Preservation Metadata Working Group

- Members:
  - Moriah Caruso - University of Washington
  - Drew Krewer - University of Houston
  - Jenny Mullins - Dartmouth College
  - Simon O’Riordan - Emory University
  - Liz Woolcott - Utah State University
  - Dave Pcolar - DPN Staff Liaison
Work Accomplished So Far
Use Case Development

To fulfill the goal of having geographically distributed copies of preservation master files, the Library decides to deposit copies of materials in an off-site dark storage environment, such as the Digital Preservation Network. Once deposited, materials cannot be changed or removed. **Files need to be packaged so that, when retrieved—whether in one, twenty or fifty years time—they can be understood, verified and used.**

Upon retrieval:

- The content and context of files must be discernable.
- There must be proof that the files are identical to the ones initially deposited.
- It must be possible to render the file's content (through current software or emulation), or relate the file to a migrated version along with documentation of the migration event.
<table>
<thead>
<tr>
<th>Property</th>
<th>UH</th>
<th>Penn State</th>
<th>Emory</th>
<th>USU</th>
<th>UW</th>
<th>Dartmouth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>dc:title</td>
<td>Title</td>
<td>Title</td>
<td>Title</td>
<td>dc.title</td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>dcterms:creator</td>
<td>Creator</td>
<td>Creator</td>
<td>Creator</td>
<td>dcterms:creator</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>dc:date</td>
<td>Date-Source</td>
<td>Date-Source</td>
<td>Date-Source</td>
<td>Date-Source</td>
<td>Date-Source</td>
</tr>
<tr>
<td>Description</td>
<td>dc:description</td>
<td>Description</td>
<td>Description</td>
<td>Description</td>
<td>dc.description</td>
<td>External-Description</td>
</tr>
<tr>
<td>Publisher</td>
<td>dcterms:publisher</td>
<td>dcterms:publisher</td>
<td>dcterms:publisher</td>
<td>dcterms:publisher</td>
<td>dcterms:publisher</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>dc:subject</td>
<td></td>
<td></td>
<td></td>
<td>dc:subject</td>
<td></td>
</tr>
<tr>
<td>Administrative Rights</td>
<td>dcterms:rights</td>
<td>Rights</td>
<td>Rights</td>
<td>Rights</td>
<td>Rights</td>
<td>Rights</td>
</tr>
<tr>
<td>Administrative Identifier</td>
<td>dcterms:identifier</td>
<td>Identifier</td>
<td>Identifier</td>
<td>Identifier</td>
<td>Identifier</td>
<td>Identifier</td>
</tr>
<tr>
<td>Administrative Institutional Identifier</td>
<td>dcterms:identifier</td>
<td>Institutional Identifier</td>
<td>Institutional Identifier</td>
<td>dcterms:identifier</td>
<td>dcterms:identifier</td>
<td>Source-Organization</td>
</tr>
<tr>
<td>Descriptive/Technical Content Type</td>
<td>Content Type</td>
<td>Content Type</td>
<td>Content Type</td>
<td>Content Type</td>
<td>Content Type</td>
<td>Content Type</td>
</tr>
<tr>
<td>Descriptive/Technical Formats</td>
<td>Format</td>
<td>Format</td>
<td>Format</td>
<td>Format</td>
<td>Format</td>
<td>Format</td>
</tr>
<tr>
<td>Descriptive/Technical Associated Items</td>
<td>Identifier-associated-identifiers</td>
<td>Identifier-associated-identifiers</td>
<td>Identifier-associated-identifiers</td>
<td>Identifier-associated-identifiers</td>
<td>Identifier-associated-identifiers</td>
<td>Identifier-associated-identifiers</td>
</tr>
<tr>
<td>Technical Filenames</td>
<td>Files</td>
<td>Files</td>
<td>Files</td>
<td>Files</td>
<td>Files</td>
<td></td>
</tr>
<tr>
<td>Technical Checksum</td>
<td>Checksum</td>
<td>Checksum</td>
<td>Checksum</td>
<td>Checksum</td>
<td>Checksum</td>
<td></td>
</tr>
<tr>
<td>Administrative Bagged by</td>
<td>Bagged by</td>
<td>Bagged by</td>
<td>Bagged by</td>
<td>Bagged by</td>
<td>Bagged by</td>
<td>Bagged by</td>
</tr>
<tr>
<td>Administrative Uploaded by</td>
<td>Uploaded by</td>
<td>Uploaded by</td>
<td>Uploaded by</td>
<td>Uploaded by</td>
<td>Uploaded by</td>
<td>Uploaded by</td>
</tr>
<tr>
<td>bag-info.txt Size</td>
<td>Size</td>
<td></td>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bae-info.txt Organization Address</td>
<td>Organization Address</td>
<td>Organization Address</td>
<td>Organization Address</td>
<td>Organization Address</td>
<td>Organization Address</td>
<td>Organization Address</td>
</tr>
</tbody>
</table>
DPN Bag-It Specification Amendment

- Member institution was creating metadata being stored as tag files.
- BagIt File Packaging Format Spec (V0.97) allows for "other tag files":
  - 2.2.4. Other Tag Files: A bag MAY contain other tag files that are not defined by this specification. Implementations SHOULD ignore the content of any unexpected tag files, except when they are listed in a tag manifest. When unexpected tag files are listed in a tag manifest, implementations MUST only treat the content of those tag files as octet streams for the purpose of checksum verification.
- Decision that optional, member-created tag files must be listed in the tag manifest to ensure checksum is recorded.
Current & Future Directions
Current Directions

University of Washington:

*A tale of “lost files and mixed blessings”*
Metadata:

Preservation gathers:

- Collection-level marc record
- EAD finding aid
- Item level metadata from access systems
- Rights data, donor agreements, supplementary info

Preservation creates: Archivematica-produced METS file with technical metadata
Current Directions

Dartmouth College

Working with what we’ve got
Future Directions

- **Case Study**
  - Finalize core record
  - Evaluate if there are additional ways we can help the membership
Future Directions

- **Focus group**
  - Representative sample of institutions
  - Analyze metadata needs, uses, and workflows
  - Gather feedback
Future Directions

- **Final Deliverables**:
  - Recommendations/best practices for DPN metadata record
  - Toolkit for partners to use
  - Paper outlining the outcomes of the project
Feedback?

- jennifer.l.mullins@dartmouth.edu
- dave@dpn.org
- neilsm@uw.edu
- ajkrewer@uh.edu
- liz.woolcott@usu.edu