Evaluating your Digital Preservation Network (DPN) metadata approach

An update from DPN’s Preservation Metadata Working Group (PMSWG)
About DPN & the PMSWG
The Digital Preservation Network

- “Dark archive”
- “Redundant and varied technical and legal infrastructure to assure the survival, ownership and management of preserved digital content”
- 60+ charter members, including public and private universities and colleges, consortia, and digital curation service providers
- Collaborative community
- Deposits started in 2016
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.
• Meets monthly through conference call, with tasks occurring in-between meetings.
Preservation Metadata Working Group

- Current Members
  - Moriah Caruso - University of Washington
  - Jennifer Mullins - Dartmouth College
  - Simon O'Riordan - Emory University
  - Andrea Payant - Utah State University
  - Erin Wolfe - University of Kansas
  - Dave Pcolar - DPN Staff Liaison
Use Case
Use Case

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.**

- **The goal of understanding the files would be met if both the content and context were discernible.**
- **The goal of verifying the files would be met if there is proof that the files are identical to the ones initially deposited.**
- **The goal of using the files would be satisfied if the file's content could be rendered (through current software or emulation), or if the file could be verifiably related to a copy in a current file format, with changes to the original well documented, as well as documentation that the file's significant properties have been maintained in the transformation.**

Meeting these goals would rely on producing metadata to be packaged with objects before offsite deposit occurs, as well as managing metadata created locally throughout the lifecycle of the object and metadata created by the storage system.
Core Record
Core Record Elements

- Title - The name of the resource being described
- Creator - The name of the person(s) or organization(s) with primary responsibility for creating the content
- Date - Date information significant to an event in the lifecycle of the original content, such as creation, publication, or issue date
- Description - Summary description of the content, such as an abstract
- Rights Statement - Information about rights held in and over the resource
- Access Rights - Information about who can access the resource or an indication of its security status
- Identifier - Unique identifier for a digital object (either a local identifier from your organization or a formal standard identifier issued and maintained by an external organization)
- Format (original) - Format of the original item represented in the digital surrogate
- Format (digital) - Format of the digital file or digital surrogate
# Core Record Crosswalks

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Simple DC</th>
<th>Qualified DC</th>
<th>MODS</th>
<th>VRA Core</th>
<th>PB Core</th>
<th>EAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>title</td>
<td>title</td>
<td>&lt;titleinfo&gt;</td>
<td>&lt;title&gt;</td>
<td>&lt;pbcoretitle&gt;</td>
<td>&lt;unititle&gt;</td>
</tr>
<tr>
<td>Creator</td>
<td>creator</td>
<td>creator</td>
<td>&lt;name&gt; with optional attributes</td>
<td>&lt;agent role=&quot;&quot;&gt; with &lt;agent name=&quot;&quot;&gt;</td>
<td>&lt;pbcreator&gt;</td>
<td>&lt;name&gt;</td>
</tr>
<tr>
<td>Date</td>
<td>date</td>
<td>dateSubmitted issued created</td>
<td>&lt;originInfo&gt;</td>
<td>&lt;date&gt;</td>
<td>&lt;pbcoreAssetDate&gt;</td>
<td>&lt;publicationstmt/date&gt;</td>
</tr>
</tbody>
</table>
Focus Groups
Five Questions

1. Tell us a little bit about yourself, what types of metadata schemas you use, and what types of content you anticipate depositing in DPN.

1. What roadblocks have you encountered that have had a direct impact on your ability to deposit content into DPN?

1. Have you thought about situations that could prompt you to retrieve content from DPN? What would your institution need to successfully restore this content?

1. How are you deciding what to put into DPN? What policies or philosophies do you have around this selection process?

1. Is there anything additional you would like to contribute to the conversation at this time? Are there things you were hoping to get out of the call that were covered?
Focus Groups

Repositories and Tools

Diversity!

Fedora, Rosetta, dSpace, CONTENTdm, local solutions

Archivematica, FITS, bagit

AP Trust, DuraCloud Vault
Focus Groups

Metadata

Diversity!

PREMIS, FITS

Dublin Core & MODS

Associating metadata with AIPs: together or apart?

Is descriptive metadata preservation metadata?
Focus Groups

Roadblocks

5TB = Too much or not enough?
First time AIP makers=What workflow???
Large files or large piles
Focus Groups

Prioritization

Campus mandates
Valuable, irreplaceable, unique
Low hanging fruit
Focus Groups

Retrieval

Disasters
Organizational failure
Metadata considerations
- Authenticity and Integrity
- Contextual information
- Significant Properties
Metadata considerations for DPN deposits

DPN-created metadata

DuraCloud Vault-created metadata

Recommended core record
Evaluating your DPN metadata approach
1. What information is needed to understand and contextualize an object?

Descriptive metadata

Structural metadata
2.

What information is needed to understand and contextualize a collection?

- Descriptive metadata
- Structural metadata
3.

How do I connect or relate objects to a collection? What are some different approaches?

Objects in a collection are deposited together

Individual items are deposited singly

Both collections and individual items are deposited, depending on...
4.

How are versions connected or related to one another?

PREMIS

...any other ideas?
5.

How do I ensure that metadata records are connected to associated objects and collections?

All metadata resides in one record.

Metadata is found in any number of records.
6. How do I ensure the authenticity of an object or collection?

Via the bag manifest/checksum created by DPN

Generate checksums before ingesting objects into DPN

Chain of custody documentation
7.

How do I distinguish original objects from derivatives?

PREMIS events & relationships
File & folder naming conventions
8. If data must be migrated on retrieval, how do I ensure that the essential characteristics of the original are known?

- Policy determination of essential characteristics
- Development or adaptation of migration pathways
- Retention of the original and migrated object
- Documentation of the original with metadata
9.

What about formats whose essential characteristics I might see as challenging to capture or understand?

- Complex objects
- Objects created with proprietary software
- Objects created with obsolete software
Conclusions
Thanks! Questions!