

Preservation Planning in the Consortial Context

*Updates from the **DPN** Preservation Metadata Standards Working Group*

Moriah Caruso, Preservation Librarian, University of Washington

Drew Krewer, Digitization Services Coordinator, University of Houston

Jennifer Mullins, Digital Preservation Librarian, Dartmouth College

Liz Woolcott, Head of Cataloging and Metadata Services, Utah State University

Outline

1. About DPN
2. Preservation Metadata Working Group- who, what, why, when, how
3. Work accomplished so far
4. Challenges encountered
5. Future plans
6. Invitation for feedback

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, consortiums, and digital curation service providers
- Collaborative community
- Members are just starting to deposit content



Preservation Metadata 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

- Members
 - Kevin Comerford - University of New Mexico
 - Drew Krewer - University of Houston
 - Jonathan Markow - DuraSpace
 - Jenny Mullins - Dartmouth College
 - Moriah Caruso - University of Washington
 - Simon O'Riordan - Emory University
 - Liz Woolcott - Utah State University
 - Dave Pcolar - DPN Staff Liaison

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 discernable.**
- **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.

		UH	Penn State	Emory	USU	UW	Dartmouth
Descriptive	Title	dc:title	Title	Title	Title	dc.title	
Descriptive	Creator	dc:creator			Creator	dc.creator	
Descriptive	Date	dc:date		Date-Source	Date-Source; Date-Digitized; Date-Ba	dc.date; dc.date.acessioned	
Descriptive	Description	dc:description	Description		Description	dc.description	External-Descriptio
Descriptive	Publisher	dc:publisher					
Descriptive	Subject					dc.subject	
Administrative	Rights	dc:rights	Rights	Rights/Access statement	Rights	dc:rights	
Administrative	Access Rights	dcterms:accessRights	Access Rights	Rights/Access statement	Access Rights		
Administrative	Identifier		Identifier	Identifier	Identifer-bagLevel	dc.identifier.uri	External-Identifier
Administrative	Institutional Identifier		Institutional Identifier		Institutional Identifier		Source-Organization
Descriptive/Administrative	Institution		Institution	Institution	Institution		
Descriptive/Technical	Content Type			Content Type	Type		
Descriptive/Technical	Formats				Format	mimetype/format	
Descriptive/Technical	Format-Extent				Format-Extent		
Descriptive/Administrative	Associated Items				Identifier-associatedIdentifiers		
Technical	Filenames				Filenames		
Technical	Checksum				Checksum		
Descriptive/Administrative	Relation-IsPartOf				Relation-IsPartOf		
Administrative	Bagged by				Bagged by		
Administrative	Uploaded by				Uploaded by		
bag-info.txt	Size					Size	
bag-info.txt	Organization Address						Organization-Addre

The goal of understanding the files would be met if both the content and context were discernable = **CONTENT, CONTEXT**

The goal of verifying the files would be met if there is proof that the files are identical to the ones initially deposited = **FIXITY**

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 = **USABILITY**

	Use Case
Title	Content
Creator	Content
Date	Content
Description	Content
Publisher	
Subject	
Rights	Content, Context
Access Rights	Content, Context
Identifier	Context, Usability
Institutional Identifier	Context
Institution	Context
Content Type	
Formats	
Format-Extent	
Associated Items	
Filenames	
Checksum	Fixity
Relation-IsPartOf	Context, Usability

Significant Properties	Usability
Relationship	Usability
Event	Usability
Environment	Usability

DPN BagIt 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.

Challenges encountered

- Defining (and refining!) the scope of the group's work
- What is preservation metadata?
- Recommended vs. required



Challenges encountered



- DPN is a new organization with
 - Very few depositors
 - Widely varying member group
 - Different sizes
 - Different technical infrastructures
 - At different places on the digital preservation continuum
 - Growing

Future plans

Case Study

- Representative sample of institutions
- Analyze metadata needs, uses, and workflows
- Gather feedback

From the results, develop :

- Recommendations/best practices for DPN metadata record
- Toolkit for partners to use



Tell us what you think

jennifer.l.mullins@dartmouth.edu

dave@dpn.org

neilsm@uw.edu

ajkrewer@uh.edu

liz.woolcott@usu.edu