Digital Preservation

ATALM Post-Conference SHN Workshop October 2016

Alex Merrill, Lotus Norton-Wisla
Agenda

- 2:15-3:30 Digital Preservation Overview
- 3:30-4:00 Storage Activity
- 4:00-4:30 Digital Preservation Action Plan
- 4:30-5:00 SHN wrap up, evaluations
Digital Preservation

- Long term storage and preservation of your digital files
- Create a Digital Preservation Plan
- Runs through all of your digital projects
- Can’t just “set it and forget it”
- Must be updating, researching, monitoring
Digital Preservation -

- Find others to bring into the conversation
- Fit your needs into what already exists
- Maintain communication
- Balance responsibilities
Storage

- Storage space for content
- Integrate with IT
  - What does your IT department already have set up?
- Consider types of storage
- Multiple layers
- Buy new storage every few years
Estimating Storage - Space Calculators

- Video - Digital Rebellion
- Audio - The Audio Archive
- Photograph - NAU
- Paper - simple Cornell equation
Storage Questions

• How often content is backed up?

• What types of storage devices are used, and how often storage devices are migrated?

• What risks are present in your region?
Fixity

- Definition: Stability of a digital object
- Files should remain unchanged over time
- Digital files can degrade without warning
Fixity Levels

- Simple steps: Expected file size, expected file count
- Good/Better: Higher level of detail MD5 (hash algorithm)
- Best: Use a use tool that uses SHA1 (hash algorithm)
How to check/store Fixity information

- Standalone software applications
  - MD5Summer, Fixity, Jacksum, BWF Metaedit

- Integrated into content management systems
  - Dspace, Islandora, Rosetta
When to Check Fixity

- Run a checksum tool
  - When created
  - At regular intervals (monthly, weekly)
  - During a change (transfer, recovery)
Security

• Know who has access to your files
• Have policies and/or technology in place to restrict access to appropriate people
• File System Logging (Something to discuss with IT)
Integrity Questions

• Does your department /IT already check fixity?
• Who on your staff (or other departments) has permissions to view/edit/delete digital files?
• Are regular virus checks run on computers? What about when new digital collections come in?
Meaning of “Access”

• Providing access to your community
• **Digital preservation** access = internal management of your files and INFORMATION about your files
Meaning of “Access”

- Adequately describing the digital object
- Ensuring continued access to the information within the file (metadata, file types, representation information, migration, security)
- By you, your staff, and future staff
Access - Types of Copies

- Preservation Master files
- Access copies
- Web Ready copies
Access - Managing File Types

- **Good**: Decide which file formats to use across collections, open formats, document which formats
- **Better**: Validate files, keep up on threats to formats
- **Best**: Migrate formats, emulation, have plans to deal with obsolescence
Access - Preservation Metadata

- Documentation
- Long-term usability, understanding

**Good:** Inventory
**Better:** Descriptive, Administrative, and Technical, metadata
**Best:** All preservation metadata
<table>
<thead>
<tr>
<th>Element name</th>
<th>Obligation</th>
<th>Suggested Value Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit Depth</td>
<td>Strongly recommended</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>Checksum</td>
<td>Strongly recommended</td>
<td>None (free text)</td>
</tr>
<tr>
<td>Collection Source</td>
<td>Required</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>Color Space</td>
<td>Optional, if applicable</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>Compression Degree</td>
<td>Recommended, if applicable</td>
<td>None (free text)</td>
</tr>
<tr>
<td>Compression Type</td>
<td>Strongly recommended, if applicable</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>Creation Hardware</td>
<td>Strongly recommended (digitized), Optional (born-digital)</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>Creation Software</td>
<td>Recommended (digitized); Strongly recommended (born-digital)</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>Digital Creation Date</td>
<td>Required</td>
<td>ISO 8601 Date-Time Format</td>
</tr>
<tr>
<td>Digital Creator</td>
<td>Required, when known</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>Digital Object ID</td>
<td>Required</td>
<td>None (free text)</td>
</tr>
<tr>
<td>Extent</td>
<td>Required</td>
<td>None (free text)</td>
</tr>
<tr>
<td>File Format</td>
<td>Required</td>
<td>Controlled vocabulary with Internet media types</td>
</tr>
<tr>
<td>File Location</td>
<td>Required</td>
<td>None (free text)</td>
</tr>
<tr>
<td>Local Repository ID</td>
<td>Strongly recommended</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>Original Object ID</td>
<td>Required, if applicable</td>
<td>None (free text)</td>
</tr>
<tr>
<td>Resolution</td>
<td>Required for static images</td>
<td>None (free text)</td>
</tr>
<tr>
<td>Revision Date</td>
<td>Strongly recommended, if applicable</td>
<td>ISO 8601 Date-Time Format</td>
</tr>
<tr>
<td>Revision History</td>
<td>Strongly recommended, if applicable</td>
<td>None (free text)</td>
</tr>
<tr>
<td>Rights Statement</td>
<td>Required</td>
<td>None (free text) or controlled vocabulary</td>
</tr>
<tr>
<td>Security</td>
<td>Optional</td>
<td>Controlled vocabulary</td>
</tr>
</tbody>
</table>
Access Questions

• What metadata scheme does your department use? Does it include preservation metadata?
• Do you have consistent file formats that you use for different types of files?
• Do you have preservation copy, access copy, edited copy for all files? How these are organized and saved?
Digital Preservation
Storage Activity
Storage Activity:
Storage Media
Shopping Cart Game
Project Parameters

- Our collection: 10 terabytes
- Our budget: $15,000
- Our grant period: 3 years
Types of Storage Media

- external hard drives
- flash drives
- RAID hard drive
- cloud/hosted storage

- SSD (solid state drives)
- LTO Tape
- Network Attached Storage
There is no cloud
it’s just someone else’s computer
Digital Preservation Action Plan
File Access
- Metadata
- File types
- Security
- Migration
- Access copies/versioning

File Integrity
- Fixity
- Security
- Write blockers
- Virus checks

File
- Storage
What is an “Action Plan”

- People in your organization
- Taking stock of what you already know
- Things that you want to find out
People in your Organization

- Supervisors
- IT Department or Staff
- Staff
- Other Departments
- Funders
- Partners
What you know already:

• What is your current storage setup?
• Who is involved in digital preservation already?
• What digital content you have now?
• What digital projects do you have planned?
What you need to find out or accomplish?

• Inventory of digital files
• Who else do you need to talk to?
• Funding sources
• Questions from: Storage, Integrity, and Access
• Find out backup schedule (when and how they happen, when and how they check backups)
Fill out your Action Plan

- 3 people to meet with
- 3 things you know already
- 3 questions to ask (yourself or others)
Resources

- The Signal - blog from Library of Congress
- Digital Preservation Q&A
- Digital Preservation Google Group
- COPTR Tool Grid
- Resources on the Sustainable Heritage Network