Preservation:

Ensuring Digital Materials Last!

KATIE FEARER AND ANASTASIA TARMANN
Digital Stewardship Lifecycle

Get it
Find materials, select materials, accept donations, make copies of state or federal records, create digital objects.

Check it
Make sure that materials are up to your standards for quality and description. Continually check that they stay intact.

Share it
Provide access to materials through exhibits, online collections, educational programs. Make available for research.

Save it
Have a secure, large enough place to save files, that can be systematically backed up with copies in multiple locations. Give your materials meaningful organization.

Source: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
Essential components

The 3 Essentials of Digital Preservation

Source: Sustainable Heritage Network 3 Essentials of Digital Preservation. Licensed under CC BY-NC-SA 4.0.
# Checklist of Sustainable Heritage Network’s Levels of Digital Preservation Preparedness

<table>
<thead>
<tr>
<th>Appropriate for</th>
<th>Minimum</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Little time or resources, little digital content, just getting started</td>
<td>A little more time and resources, more digital content, understand and are carrying out basic steps</td>
<td>Have funding, support, time, and have already gotten most intermediate steps in place</td>
</tr>
<tr>
<td><strong>Storage</strong></td>
<td>• 2+ copies in different geographic locations</td>
<td>• Follow 3:1-2:1 rule (3 copies, 2 types of storage media, at least 1 in different geographic location)</td>
<td>• Follow at least 3:1-2:1 plan but also ensure that at least 1 copy is in geographic area with different disaster threat</td>
</tr>
<tr>
<td></td>
<td>• Transfer data from media such as DVDs to your storage system upon receipt</td>
<td>• Monitor all storage systems and media</td>
<td>• Comprehensively plan to ensure files and metadata will stay in accessible media and systems</td>
</tr>
<tr>
<td></td>
<td>• Inventory of storage media and practices in use</td>
<td>• Plan for replacement of storage media after recommended lifespan</td>
<td>• Transfer files off of all obsolete storage media and properly dispose</td>
</tr>
<tr>
<td></td>
<td>• Basic emergency or disaster plan for when storage fails</td>
<td>• Transition away from storage media more than 10 years old</td>
<td>• Migrate from old to new media at set schedule (use manufacturer’s recommendations)</td>
</tr>
<tr>
<td></td>
<td>• Ensure use of limited set of known and open file formats</td>
<td>• Emergency or disaster plan for when storage fails</td>
<td>• Create and test detailed emergency or disaster plan</td>
</tr>
<tr>
<td><strong>Access</strong></td>
<td>• Identify who has what roles with regard to files</td>
<td>• Maintain logs of who accesses files</td>
<td>• Maintain logs of who performs what actions on files, including deletions and preservation actions</td>
</tr>
<tr>
<td></td>
<td>• Restrict ability to access to necessary people</td>
<td>• Store metadata for digital content</td>
<td>• Perform audit of logs</td>
</tr>
<tr>
<td></td>
<td>• Inventory of content</td>
<td>• Validate files against file formats with software validation tools</td>
<td>• Store metadata for digital content, including preservation metadata</td>
</tr>
<tr>
<td></td>
<td>• Inventory of file formats</td>
<td>• Monitor file format obsolescence threats</td>
<td>• Perform file format migration when formats become outdated or unsupported</td>
</tr>
<tr>
<td></td>
<td>• Encourage use of limited set of known and open file formats</td>
<td></td>
<td>• Perform emulation if migration does not produce intended results</td>
</tr>
<tr>
<td><strong>Integrity</strong></td>
<td>• Check fixity on ingest</td>
<td>• Check fixity on ingest</td>
<td>• Check fixity after any transformative act</td>
</tr>
<tr>
<td></td>
<td>• Check at least file size and count</td>
<td>• Use at least MD5 hash algorithm</td>
<td>• Check fixity of at least a sample of files at fixed intervals</td>
</tr>
<tr>
<td></td>
<td>• Use at least</td>
<td>• Use write blockers when working with original media (to prevent additional content being written to that media)</td>
<td>• Use SHA1 or SHA256 hash algorithm</td>
</tr>
<tr>
<td></td>
<td>• If fixity information not provided with content on ingest use a tool to do it and store it in a file or spreadsheet</td>
<td>• Perform virus checks on high-risk content</td>
<td>• Maintain logs of fixity information and audit on demand</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Virus check all content</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• In emergency, check fixity as determined necessary and have plan in place to replace corrupted data</td>
</tr>
</tbody>
</table>

Adapted from: Sustainable Heritage Network Levels of Digital Preservation Preparedness. Licensed under CC BY-NC-SA 4.0.
Storage

Stack of Magazines and Books, by Slick-o-bot,
Public Domain.
Initial Activities to Create a Digital Preservation Program

- Conversations and meetings with people in your department and organization
- Inventory of existing digital content
- Assess staffing needs for digital preservation activities
- Research tools, equipment, other policies that you may want to use in your organization
- Plan (and purchase if needed) storage solutions, storing multiple copies in multiple places on multiple formats
- Create plans for digital content in case of disaster or emergency
- Create plans for digital content in case of disaster or emergency
  - Know the natural or man-made disasters that might affect your region, and create plans of what to do in response
  - Also consider hardware failures, network errors, network security and external attacks, software failure, media failure or obsolescence.
- Select and identify materials for long term digital preservation.

Upon ingest and file creation

- Save master preservation file and access file
- Add to inventory

Regularly

- Backup files on a regular basis

Less Frequently

- Research new tools, equipment, or policies that you may want to use in your organization
- Update storage media

In response to disaster, emergency

- Follow your digital disaster or emergency plan - assuming you have created one as part of your overall Digital Preservation Plan
- Assess what loss or damage has occurred
- Retrieve all possible content from your backup system
Storage

Ensuring digital files are retrievable; How?

- Make and update policies and plans
- Use preservation metadata and standards
- 2+ copies of raw files, 1 separate geographic location
- Update media and storage devices, media, and formats
- Run back-ups
- Explore saving and storage options
Storage at SLAM

- NetApps – virtual smart storage, mirrored in Anchorage
- RAID – multiple hard drives, replicates files (drive assigned to Archives). Will indicate if a drive fails, no fixity checks, but data is not lost unless too many drives fail. Manual ingest. Stand alone. Danger! Only duplicates: Family Search, books, etc.
- Tape backup every three months
# Storage levels: which is right for you?

<table>
<thead>
<tr>
<th>Storage</th>
<th>Minimum</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriate for</td>
<td>Little time or resources, little digital content, just getting started</td>
<td>A little more time and resources, more digital content, understand and are carrying out basic steps</td>
<td>Have funding, support, time, and have already gotten most intermediate steps in place</td>
</tr>
<tr>
<td>• 2+ copies in different geographic locations</td>
<td>• Follow 3-2-1 rule (3 copies, 2 types of storage media, at least 1 in different geographic location)</td>
<td>• Follow at least 3-2-1 plan but also ensure that at least 1 copy is in geographic area with different disaster threat</td>
<td></td>
</tr>
<tr>
<td>• Transfer data from media such as DVDs to your storage system upon receipt</td>
<td>• Monitor all storage systems and media</td>
<td>• Comprehensive plan to ensure files and metadata will stay in accessible media and systems</td>
<td></td>
</tr>
<tr>
<td>• Inventory of storage media and practices in use</td>
<td>• Plan for replacement of storage media after recommended life span</td>
<td>• Transfer files off of all obsolete storage media and properly dispose</td>
<td></td>
</tr>
<tr>
<td>• Basic emergency or disaster plan for when storage fails</td>
<td>• Transition away from storage media more than 10 years old</td>
<td>• Migrate from old to new media at set schedule (use manufacturer's recommendations)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Emergency or disaster plan for when storage fails</td>
<td>• Create and test detailed emergency or disaster plan</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from: [Sustainable Heritage Network Levels of Digital Preservation Preparedness](https://www.sustainable-heritage.net/). Licensed under CC BY-NC-SA 4.0.
Access

Closet, licensed under CC BY-NC 4.0.
Access

- Versioning: Different types of files for different purposes: preservation (raw), prettied up, access, Web compressed
- File type: proprietary versus open
- Ensuring continued access: migration, emulation, replication
- Logs: who is accessing and doing what with what? Audit trail
- Preservation metadata: PREMIS
PREMIS = Preservation Metadata

- Supports long-term usability (remember the messy closet?)
- Supports readability
- Supports currency versus obsolescence
- Supports authenticity

## Preservation metadata functions

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Supports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checksum</td>
<td>Secures against modification</td>
</tr>
<tr>
<td>Media: dates, iterations, types, length of storage</td>
<td>Media management: currency versus obsolescence</td>
</tr>
<tr>
<td>Original file formats, software, hardware</td>
<td>Continued access to files (track preservation strategies: migration, emulation)</td>
</tr>
<tr>
<td>Provenance</td>
<td>Authenticity (changes in rendering, etc.)</td>
</tr>
</tbody>
</table>

C. What is the PMDO?

The Preservation Metadata for Digital Objects (PMDO) element set addresses the metadata needs for the preservation and long-term management of digital files. It neither addresses descriptive metadata elements such as those in Dublin Core, nor structural metadata such as TEI (Text Encoding Initiative) or EAD (Encoded Archival Description). Instead, the PMDO element set is meant to collect metadata beyond what is already being created using other standards.

Originally, the PMDO was developed to assist institutions with the collection of preservation metadata for scanned images and text. This updated version of the PMDO addresses what are commonly referred to as “born-digital” files, or files originally created on electronic devices, such as word processing and spreadsheet files, digital photographs, websites, and email.

The PMDO, like other element sets, is comprised of a series of elements into which data values are input. In that way, it is similar to Dublin Core. The PMDO elements are different from those in other metadata element sets and require implementers to create different data values.
## NC PMDO: North Carolina Preservation Metadata for Digital Objects

<table>
<thead>
<tr>
<th>Element name</th>
<th>Obligation</th>
<th>Suggested Value Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Bit Depth</td>
<td>Strongly recommended</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>2  Checksum</td>
<td>Strongly recommended</td>
<td>None (free text)</td>
</tr>
<tr>
<td>3  Collection Source</td>
<td>Required</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>4  Color Space</td>
<td>Optional</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>5  Compression Degree</td>
<td>Recommended, if applicable</td>
<td>None (free text)</td>
</tr>
<tr>
<td>6  Compression Type</td>
<td>Strongly recommended, if applicable</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>7  Creation Hardware</td>
<td>Strongly recommended (digitized); Optional (born-digital)</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>8  Creation Software</td>
<td>Recommended (digitized); Strongly recommended (born-digital)</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>9  Digital Creation Date</td>
<td>Required</td>
<td>ISO 8601 Date-Time Format</td>
</tr>
<tr>
<td>10 Digital Creator</td>
<td>Required, when known</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>11 Digital Object ID</td>
<td>Required</td>
<td>None (free text)</td>
</tr>
<tr>
<td>12 Extent</td>
<td>Required</td>
<td>None (free text)</td>
</tr>
<tr>
<td>13 File Format</td>
<td>Required</td>
<td>Controlled vocabulary with Internet media types</td>
</tr>
<tr>
<td>14 File Location</td>
<td>Required</td>
<td>None (free text)</td>
</tr>
<tr>
<td>15 Local Repository ID</td>
<td>Strongly recommended</td>
<td>Controlled vocabulary</td>
</tr>
<tr>
<td>16 Original Object ID</td>
<td>Required, if applicable</td>
<td>None (free text)</td>
</tr>
<tr>
<td>17 Resolution</td>
<td>Required for static images</td>
<td>None (free text)</td>
</tr>
<tr>
<td>18 Revision Date</td>
<td>Strongly recommended, if applicable</td>
<td>ISO 8601 Date-Time Format</td>
</tr>
<tr>
<td>19 Revision History</td>
<td>Strongly recommended, if applicable</td>
<td>None (free text)</td>
</tr>
<tr>
<td>20 Rights Statement</td>
<td>Required</td>
<td>None (free text) or controlled vocabulary</td>
</tr>
<tr>
<td>21 Security</td>
<td>Optional</td>
<td>Controlled vocabulary</td>
</tr>
</tbody>
</table>
Example of an element

1. Bit Depth

**Obligation:** Strongly recommended

**Definition:** Number of bits (1s and 0s) used to represent the smallest unit of information (such as a pixel) in an audio, video, or graphic data file.

*Digital audio/video* – Describes the number of bits of information recorded for each sample. A sample refers to a value or set of values at a point in time and/or space.

*Digital image* – The number of bits used to represent the color of a single pixel.

**Input guidelines:** Create a controlled vocabulary. Standard values may include:

<table>
<thead>
<tr>
<th>Digital file type</th>
<th>Bit depth</th>
<th>This means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-tonal</td>
<td>1 bit</td>
<td>Each pixel is either black or white</td>
</tr>
<tr>
<td>Grayscale</td>
<td>8 bit</td>
<td>Each pixel can be 1 of 256 shades of gray</td>
</tr>
<tr>
<td>Color</td>
<td>8 bit</td>
<td>Each pixel can be 1 of 256 shades of color</td>
</tr>
<tr>
<td>Color</td>
<td>24 bit</td>
<td>Each pixel can be 1 of 16.8 million shades of color (sometimes referred to as &quot;millions of colors&quot;). This is considered &quot;true&quot; color, and most color files are now created at 24 bit.</td>
</tr>
<tr>
<td>Audio</td>
<td>16 bit</td>
<td>Signal-to-noise ratio maximum of 96 dB</td>
</tr>
<tr>
<td>Audio</td>
<td>24 bit</td>
<td>Signal-to-noise ratio maximum of 144 dB</td>
</tr>
</tbody>
</table>

# Access levels: which is right for you?

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appropriate for</strong></td>
<td>Little time or resources, little digital content, just getting started</td>
<td>A little more time and resources, more digital content, understand and are carrying out basic steps</td>
<td>Have funding, support, time, and have already gotten most intermediate steps in place</td>
</tr>
<tr>
<td><strong>Access</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Identify who has what roles with regard to files</td>
<td>• Maintain logs of who accesses files</td>
<td>• Maintain logs of who performs what actions on files, including deletions and preservation actions</td>
</tr>
<tr>
<td></td>
<td>• Restrict ability to access to necessary people</td>
<td>• Store metadata for digital content</td>
<td>• Perform audit of logs</td>
</tr>
<tr>
<td></td>
<td>• Inventory of content</td>
<td>• Validate files against file formats with software validation tools</td>
<td>• Store metadata for digital content, including preservation metadata</td>
</tr>
<tr>
<td></td>
<td>• Inventory of file formats</td>
<td>• Monitor file format obsolescence threats</td>
<td>• Perform file format migration when formats become outdated or unsupported</td>
</tr>
<tr>
<td></td>
<td>• Encourage use of limited set of known and open file formats</td>
<td></td>
<td>• Perform emulation if migration does not produce intended results</td>
</tr>
</tbody>
</table>

Adapted from: [Sustainable Heritage Network Levels of Digital Preservation Preparedness](https://www.sustainableheritagenet.org/main/users). Licensed under CC BY-NC-SA 4.0.
Integrity

Coat Closet by Lisa Yarost, licensed under CC BY 2.0
Integrity: what and how

- Ensuring digital files remain **viable, usable, safe, unchanged**

- How to do it:
  - Check fixity
  - Ensure security
  - Use write blockers
  - Run virus scans

Source: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
Check fixity

Stability of a digital object

How to do it

- At least: check file size and count
- Better: MD5 hash algorithm
- Best: SHA1 or SHA256 hash algorithm

Source: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
Checksums

- Need program that uses a hash algorithm
- Unique string of letters and numbers associated with the file
- Checksum changes as file changes
MD5 vs SHA256 checksums

**MD5**

<table>
<thead>
<tr>
<th>MD5 Checksum</th>
<th>File Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>5bd41ba0f43300058aa670397a0245d9</td>
<td>C:\Users\katie\Documents\AKLA\akla-j</td>
</tr>
<tr>
<td>735a2373e221e2c4860d5a73b2a78284</td>
<td>C:\Users\katie\Documents\AKLA\akla-j</td>
</tr>
<tr>
<td>d7ec2b9bc63c8fc55a02f25854411513</td>
<td>C:\Users\katie\Documents\AKLA\Data slu</td>
</tr>
<tr>
<td>c857edab46026914f0960edfa435e04b</td>
<td>C:\Users\katie\Documents\AKLA\Data wit</td>
</tr>
</tbody>
</table>

**SHA256**

<table>
<thead>
<tr>
<th>SHA256 Checksum</th>
<th>File Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>eytta111113\TS\istency\f536837841e6250f38c8e8b83c2f599e123\da24848/9</td>
<td>C:\Users\katie\Documents\AKLA\akla-j</td>
</tr>
<tr>
<td>f66007073c3ab79478277064a07091622b09a14349d3f413c5eb3e736e9286</td>
<td>C:\Users\katie\Documents\AKLA\akla-j</td>
</tr>
<tr>
<td>e15f845be6aca691415b7b068bfbbf8079239f980950c79ab3acb9518a75384</td>
<td>C:\Users\katie\Documents\AKLA\Data slu</td>
</tr>
<tr>
<td>bc9566772e2e7c87a2ec4ebdc3bf20762dfd09ba7ca056ecf394bb1db861dbc9</td>
<td>C:\Users\katie\Documents\AKLA\Data wit</td>
</tr>
<tr>
<td>faced43d1b13a22224bf3e80903d230656a1d32b335a13e4f7bf1f41f5a3b7d45</td>
<td>C:\Users\katie\Documents\AKLA\Data wi</td>
</tr>
</tbody>
</table>
Tools to check fixity

- Standalone software, including open source
  - MD5Summer, Fixity, Jacksum, BWF Metaedit

- Integrated into complete or partial preservation systems
  - Preservica, DSpace, Islandora, Duracloud, Rosetta

Source: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
Category: Fixity

See also
- DP Q&A questions tagged with 'fixity'
- DP Q&A: What tools do you use for the ongoing monitoring of checksums?

Function definition: Tools that support the verification of file fixity, typically through the generation and validation of checksum-based manifests.

Pages in category "Fixity"

The following 18 pages are in this category, out of 18 total.

A
- ACE (Audit Control Environment)

B
- Bagger
- BagIt Transfer Utilities
- BIL (BagIt Library)
- BitCurator

C
- Cksum Unix command

F
- File Analyzer and Metadata Harvester V2
- FileVerifier++
- Fxi
- Fxidy

M
- Md5deep and hashdeep
- Md5sum Unix command

M cont.
- Md5summer

N
- NARA File Analyzer and Metadata Harvester

P
- Python checkm package

R
- Rhash

S
- SAFE Archive Audit System
MD5Summer in action
(video from UK Data Service)

How to use MD5summer to check the integrity of your research files

Scott Summers

Video tutorial
March 2016
MD5Summer

Source: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
Fixity (the program)

- Scan up to 7 directories at once, any file type
- Creates manifest of file path, checksums, index values
- Can use MD5 or SHA526 checksum algorithm
- Runs comparative analysis monthly, weekly, daily
- Creates report on file integrity and attendance
- Emails reports
Fixity interface
A fixity email


from fearerke@gmail.com

to me

Fixity report
Project name Test
Algorithm used sha256
Date 2019-03-10
Time Elapsed 0 hrs 0 min 11 seconds
Total Files 21
Confirmed Files 19
Moved or Renamed Files 0
New Files 0
Changed Files 2
Removed Files 1
Error: Index Value Equals Zero 2

f

fixity_2019-03-10....
Logs saved by Fixity in scanned directory
## Fixity report and history files

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixity report</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project name</td>
<td>Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algorithm used</td>
<td>sha256</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Elapsed</td>
<td>0 hrs 0 min 0 seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Files</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confirmed Files</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moved or Renamed Files</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Files</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Changed Files</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removed Files</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error: Index Value Equals Zero</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Confirmed File:**
- C:\Users\Katia\Documents\AKLA\akla-j minutes 1-18-19.docx
- C:\Users\Katia\Documents\AKLA\akla-j minutes 12-7-18.docx
- C:\Users\Katia\Documents\AKLA\akla-j minutes 1-18-19.doc
- C:\Users\Katia\Documents\AKLA\Data without Borders 2.pptx
- C:\Users\Katia\Documents\AKLA\Data without Borders 3.pptx
- C:\Users\Katia\Documents\AKLA\Endangered data.docx
- C:\Users\Katia\Documents\AKLA\Endangered data.docx
- C:\Users\Katia\Documents\AKLA\Website Archaeology.pptx
- C:\Users\Katia\Documents\AKLA\Website Archaeology.ppt
- C:\Users\Katia\Documents\AKLA\Website Archaeology.pptx
- C:\Users\Katia\Documents\AKLA\Website Archaeology.pptx
- C:\Users\Katia\Documents\AKLA\Site slides.docx
- C:\Users\Katia\Documents\AKLA\Site slides.docx
- C:\Users\Katia\Documents\AKLA\Site slides.docx
- C:\Users\Katia\Documents\AKLA\new Site slides.docx
- C:\Users\Katia\Documents\AKLA\akla.j minutes 1-18-19.docx
- C:\Users\Katia\Documents\AKLA\new Site slides.docx
- C:\Users\Katia\Documents\AKLA\akla.j minutes 13-7-18.docx

**Error: Index Value Equals Zero:**
- C:\Users\Katia\Documents\AKLA\akla-j minutes 1-18-19.docx changed to C:\Users\Katia\Documents\AKLA\akla-j minutes 12-7-18.docx
- C:\Users\Katia\Documents\AKLA\akla-j minutes 12-7-18.docx changed to C:\Users\Katia\Documents\AKLA\akla-j minutes 1-18-19.docx
- C:\Users\Katia\Documents\AKLA\akla-j minutes 1-18-19.docx changed to C:\Users\Katia\Documents\AKLA\akla-j minutes 12-7-18.docx

**Error: Index Value Equals Zero:**
- C:\Users\Katia\Documents\AKLA\akla-j minutes 1-18-19.docx changed to C:\Users\Katia\Documents\AKLA\akla-j minutes 12-7-18.docx
- C:\Users\Katia\Documents\AKLA\akla-j minutes 12-7-18.docx changed to C:\Users\Katia\Documents\AKLA\akla-j minutes 1-18-19.docx
- C:\Users\Katia\Documents\AKLA\akla-j minutes 1-18-19.docx changed to C:\Users\Katia\Documents\AKLA\akla-j minutes 12-7-18.docx
Checksum stored in a database

<table>
<thead>
<tr>
<th>dc.date.issued</th>
<th>2007-11-06</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.identifier.uri</td>
<td><a href="http://hdl.handle.net/2376/1116">http://hdl.handle.net/2376/1116</a></td>
</tr>
<tr>
<td>dc.description.abstract</td>
<td>This presentation of images from the Influenza Epidemic of 1918 at WSU digital collection <a href="http://content.wsulibs.wsu.edu/cdm4/index_wsu_flu.php">http://content.wsulibs.wsu.edu/cdm4/index_wsu_flu.php</a> was shown in Beasley Coliseum prior to a talk by Gina Kolata. Beverly Makhani suggested the presentation.</td>
</tr>
<tr>
<td>dc.description.provenance</td>
<td>Submitted by Trevor Bond (<a href="mailto:tjbond@wsu.edu">tjbond@wsu.edu</a>) on 2007-11-10 No. of bitstreams: 1 110507wsufluPPTTrevor.ppt: 7476736 bytes, checksum: 32ce9858ce45a3ae9944f78609e810d0 (MD5)</td>
</tr>
<tr>
<td>dc.description.provenance</td>
<td>Made available in DSpace on 2007-11-10 (GMT), No. of bitstreams: 1 110507wsufluPPTTrevor.ppt: 7476736 bytes, checksum: 32ce9858ce45a3ae9944f78609e810d0 (MD5) Previous issue date: 2007-11-06</td>
</tr>
<tr>
<td>dc.language.iso</td>
<td>en_US</td>
</tr>
<tr>
<td>dc.subject</td>
<td>1918 Influenza</td>
</tr>
<tr>
<td>dc.subject</td>
<td>Common Reading</td>
</tr>
<tr>
<td>dc.title</td>
<td>The 1918 Influenza Pandemic at WSU</td>
</tr>
<tr>
<td>dc.type</td>
<td>Presentation</td>
</tr>
</tbody>
</table>

Source: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
When you find an error

- Two files with same name, each with different checksum
- Which is correct?
- 3-2-1 may make it easier to decide
When to check fixity

- Ingest
- Regular intervals (at least monthly)
- During a change (transfer, recovery)

Source: Sustainable Heritage Network.Licensed under CC BY-NC-SA 4.0.
Ensure security

- Know who has access to your files
- Have policies and/or technology in place to restrict access
- File system logging

Source: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
Integrity is also about quality, accuracy

- Digitize once, creating highest quality preservation copy
- Check accuracy of metadata
- Have multiple people involved

Source: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
What to do

- Inventory digital content and how you get it
- Inventory and/or revise policies and procedures
  - Ingest, digitization, transfer, quality control, access, fixity and virus checking
- Acquire and use tools
- Revise and/or establish disaster plan

Adapted from: Sustainable Heritage Network. Licensed under CC BY-NC-SA 4.0.
Systems
System management

- Locally managed (software installed in a local environment) – requires labor and technical expertise
- Hosted and maintained by 3rd party – more expensive, less labor/expertise
- Collaborative
Key factor in tool choice: infrastructure

- Equipment
- Policies, priorities, practices
- Partnerships you can leverage
- People – how many, programming skills and software expertise
- Your other systems
- Budget
### Assessing a preservation system

<table>
<thead>
<tr>
<th>Ingest process</th>
<th>Interoperability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods: bit-level, migration, emulation</td>
<td>Dark &amp; open archive</td>
</tr>
<tr>
<td>Formats supported</td>
<td>Cost</td>
</tr>
<tr>
<td># copies, where stored</td>
<td>Administration &amp; reporting</td>
</tr>
<tr>
<td>Metadata schemas</td>
<td>Longevity of system &amp; host</td>
</tr>
<tr>
<td>Fixity, authenticity, versions</td>
<td>Mechanism for exit</td>
</tr>
</tbody>
</table>
Key factor in tool choice: what you preserve and who uses it

- Images, documents, websites, datasets, software
- Born digital vs digitized
- Owned vs licensed
- Restricted vs freely available
# Create your own evaluation matrix

**N.C. State Library/Archives**

| System allows for assigning of user roles with varying levels of security access. | Tests system’s ability to provide, edit, and delete options for user management. | 1. Create 3 user accounts (Alpha, Beta, Gamma) within system.  
2. Assign full administrative rights to user Alpha.  
3. As userAlpha, create user Beta, granting full administrative rights.  
4. As user Beta, create user Gamma, assigning only one, low level privilege.  
5. As user Alpha, remove user Gamma from the system.  
SCORE: Record a score of 1 for each successful step. Record a 0 for each unsuccessful step. Return the sum of scores divided by 5. | All five test steps passed using web console. | 1 | nstln.or
| System records all actions performed by users within the system. | Provides documentation of system’s ability to document all significant user actions and make such documentation accessible to appropriate users. | 1. Provide citation of acceptable statement by system provider regarding comprehensive logging of user actions.  
2. Provide citation of acceptable statement by system provider regarding user permission based access to said logs.  
SCORE: Record a score of 1 for each successful step. Record a 0 for each unsuccessful step. Return the sum of scores divided by 2. | Lorenzo (issue ... | 0.6 | nstln.or
| System meets NCDIT security requirements and conforms to best practices to prevent data loss or malicious attack on system or NCDIT servers. | Provides documentation of system security measures and NCDIT acceptance of system security features. | 1. Provide citation of acceptable statement by system provider regarding data loss prevention practices.  
2. Provide citation of acceptable statement by system provider regarding data breach prevention practices.  
3. Provide citation of statement by appropriate NCDIT staff as to acceptance of system security features.  
SCORE: Record a score of 1 for each citation provided for statements deemed acceptable by NCDIT and NCDIT staff. Record a score of 0 if no citation is provided or statement is not acceptable. Return the sum of scores divided by 2. | Lorenzo (issue ... | 0.667 | nstln.or
| Test ingested | Tests system’s ability to create SIF package from existing BagIt bags. | 1. Using system, create SIF by importing bag sample located at /next/files/2/2/bag.  
Note: The bag must be imported as a Bag It bag as opposed to simply creating a SIF from the content within the bag helper.  
2. Verify SIF checksums match BagIt checksums.  
SCORE: Record a score of 1 for each successful step. Record a 0 for each unsuccessful step. Return the sum of scores divided by 2. | Lorenzo (issue ... | 0.5 | nstln.or
| System can ingest unstructured files without metadata. | Tests system’s ability to create SIF from data files without concerning metadata. | 1. Using system, create SIF by from the non-described content located at /next/files/2/2/ | Lorenzo (issue ... | 1 | nstln.or
Example: Preservica

- Supports many life cycle functions including ingest, storage, data management, migration, fixity checks, public access
- About $12,000 for first TB, $550 - $1,450 for each additional TB (Amazon Glacier vs. Amazon S3)
- Used by Kentucky Department of Libraries and Archives
Example: Duracloud

- Storage (Amazon), web interface & tools to ingest, check, sync, restore
- Integrates with Archive-It, Archivematica, DSpace
- Basic – 1 copy, up to 5 TB, $1,235 plus $700 per TB
- Plus – 2 or more copies, up to 5 TB, $1,235 plus $825 per TB
- Additional enterprise options for more storage, shared access
Contact Us!

Katie Fearer, Documents Librarian: kathleen.fearer@alaska.gov

Anastasia Tarmann, Historical Collections Librarian &

AK Digital Newspaper Project Director: anastasia.tarmann@alaska.gov

library.alaska.gov
Share, adapt, reuse

- Many of the materials in this presentation were adapted from the Sustainable Heritage Network.

- Unless otherwise noted, each of these slides are available under a Creative Commons Attribution Non-Commercial Share Alike license.
Resources

- Sustainable Heritage Network, Module 3: Save It
- Understanding PREMIS, Library of Congress
- Recommended Formats Statement, Library of Congress
- COPTR (Community Owned Digital Preservation Tool Registry)
- System Evaluation Matrix, State Library of North Carolina