Initial Steps on a Digital Preservation Policy

Brittany Hayes, University of Akron Archives
Zoe Orcutt, University of Akron Archives
Sidney Gao, University of Cincinnati Libraries
James Van Mil, University of Cincinnati Libraries

https://bit.ly/2N1hQS2
UC Libraries

Sidney Gao
- Digital Imaging Coordinator
- Background in management and workflows for digitization

James Van Mil
- Digital Projects & Preservation Librarian
- Work history outside of digital collections; was acquisitions/e-resources librarian and software developer.

UA Archives

Brittany Hayes
- Library Cataloging Specialist/Electronic Resources
- Assists with audio-visual digitization in the University Archives

Zoe Orcutt
- Library Research Assistant at Archival Services
  - Researcher for the University of Akron’s Sesquicentennial
Our Repositories: What are we preserving?

UC Libraries

- DRC (antique DSpace installation)
  - Historical and Archival Records
- Scholar@UC (Samvera)
  - Faculty-owned
- Luna Image repository
  - Art, but also image-based archives content
- Other content in storage
- Current digitization workflow

UA Archives

- The archives and special collections of the university
- Materials related to:
  - The University of Akron
  - Akron and Summit County
  - The rubber industry
  - Goodyear Tire & Rubber Company
  - Lighter-than-air flight
  - Ohio canals
  - The book and print culture
  - The B-26 Marauder and the 9th Air Force
What is digital preservation?

Maintenance of digital objects to ensure future accessibility and use.
Three legged stool of digital preservation

Know your challenges

Mold your policy
### Levels of Digital Preservation

<table>
<thead>
<tr>
<th>Functional Area</th>
<th>Level 1 (Know your content)</th>
<th>Level 2 (Protect your content)</th>
<th>Level 3 (Monitor your content)</th>
<th>Level 4 (Sustain your content)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>Have two complete copies in separate locations</td>
<td>Have three complete copies with at least one copy in a separate geographic location</td>
<td>Have at least one copy in a geographic location with a different disaster threat than the other copies</td>
<td>Have at least three copies in geographic locations, each with a different disaster threat</td>
</tr>
<tr>
<td></td>
<td>Document all storage media where content is stored</td>
<td>Document storage and storage media indicating the resources and dependencies they require to function</td>
<td>Have at least one copy on a different storage media type</td>
<td>Maximize storage diversification to avoid single points of failure</td>
</tr>
<tr>
<td></td>
<td>Put content into stable storage</td>
<td>Track the obsolescence of storage and media</td>
<td></td>
<td>Have a plan and execute actions to address obsolescence of storage hardware, software, and media</td>
</tr>
<tr>
<td>Integrity</td>
<td>Verify integrity information if it has been provided with the content</td>
<td>Verify integrity information when moving or copying content</td>
<td>Verify integrity information of content at fixed intervals</td>
<td>Verify integrity information in response to specific events or activities</td>
</tr>
<tr>
<td></td>
<td>Generate integrity information if not provided with the content</td>
<td>Use write-blockers when working with original media</td>
<td>Document integrity information verification processes and outcomes</td>
<td>Replace or repair corrupted content as necessary</td>
</tr>
<tr>
<td></td>
<td>Virus check all content; isolate content for quarantine as needed</td>
<td>Back up integrity information and store copy in a separate location from the content</td>
<td>Perform audit of integrity information on demand</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>Determine the human and software agents that should be authorized to read, write, move, and delete content</td>
<td>Document the human and software agents authorized to read, write, move, and delete content and apply these</td>
<td>Maintain logs and identify the human and software agents that performed actions on content</td>
<td>Perform periodic review of actions/access logs</td>
</tr>
<tr>
<td>Metadata</td>
<td>Create inventory of content, also documenting current storage locations</td>
<td>Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)</td>
<td>Determine what metadata standards to apply</td>
<td>Record preservation actions associated with content and when those actions occur</td>
</tr>
<tr>
<td></td>
<td>Backup inventory and store at least one copy separately from content</td>
<td>Find and fill gaps in your metadata to meet those standards</td>
<td>Implement metadata standards chosen</td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>Document file formats and other essential content characteristics including how and when these were identified</td>
<td>Verify file formats and other essential content characteristics</td>
<td>Monitor for obsolescence, and changes in technologies on which content is dependent</td>
<td>Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed</td>
</tr>
<tr>
<td></td>
<td>Build relationships with content creators to encourage sustainable file choices</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Level</th>
<th>Electronic Records Survey Capability Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>The organization has little or no capability or resources to collect and analyze information about the volume, location, media, format types, and lifecycle management requirements for electronic records.</td>
</tr>
<tr>
<td>1</td>
<td>The organization uses existing retention schedules to identify electronic records of permanent historical, fiscal, and legal value in the custody of records producers. It may also conduct ad hoc, one-time interviews and surveys to identify other electronic records of permanent historical, fiscal, and legal value.</td>
</tr>
<tr>
<td>2</td>
<td>The organization uses systematic interviews, surveys, and retrospective analysis of existing retention schedules to identify electronic records of permanent historical, fiscal, and legal value in the custody of select records producers. This effort may be enhanced by focusing on identified “at risk” electronic records.</td>
</tr>
<tr>
<td>3</td>
<td>The organization supplements analysis of “at risk” electronic records through collection of information about the volume and location, media and format types (preservation ready and near-preservation ready) of permanent electronic records in the custody of records producers.</td>
</tr>
<tr>
<td>4</td>
<td>The organization has identified and categorized all preservation ready, near-preservation ready, and legacy permanent electronic records in the custody of all records producers.</td>
</tr>
</tbody>
</table>

ISO 14721 Conformance

Common Challenges

Limited resources & staffing

Legacy content

Deferring the Work
Limited Resources & Staffing

University of Cincinnati Libraries

- Intentionally slow pace accommodates limited staffing
- Progressive approach that continuously builds on existing infrastructure
- Compensates with documentation, and by sacrificing speediness

University of Akron Archives

- Continuing self-education in digital preservation
  - Research and attend workshops
- Strategically work around budget restrictions
- Creating a team of staff willing/able to assist
Legacy Content

University of Cincinnati Libraries

- Developing collection assessment strategies
  - Relationships with stakeholders
  - Documenting collection goals and technical debt
- Balance investment in maintenance of established collections vs investment in new content

University of Akron Archives

- Identify all legacy content
  - We are at this stage
- Keep a tracking sheet
- Create a plan of action for incoming, obsolete materials
- Request that incoming materials be in a standardized file format as outlined by the DPCMM
Deferring the Work

University of Cincinnati Libraries

- Set small, achievable goals for your level of staffing and resources
- Standing meetings between core collaborators relieves the emotional labor of planning
- Remember that digital preservation work is often front-loaded
- Avoid deferring responsibility (do not run by committee)

University of Akron Archives

- We do have meetings every other week
  - It is slow paced because we are still doing the assessment.
- With the current pandemic we are doing a lot of research that we wouldn’t have been doing otherwise.
Unique Challenges

UC Libraries
- Accessibility

UA Archives
- Finding the right Digital Asset Management (DAM) system
UCL Accessibility *is* Digital Preservation

Culture of Accessibility Compliance

“PDF/A considered harmful for digital preservation”

Formats and information accessibility
UA - How do we find the right DAMS for our content?

- Searching for the right Digital Asset Management/Archival Management system
  - Budgeting obstacles
  - Administrative support
- CONTENTdm
  - Institutional repository
- Remote accessibility
  - Patrons and users can access content without physically coming to facility
  - We’ve seen an increase of usage of our DAM with the coronavirus pandemic
    - Digital preservation more relevant with current pandemic
Zippy with Yogi Bear

One black and white photographic print of the Zippy the kangaroo mascot posing with Yogi Bear on the field at a University of Akron football game, probably at the Rubber Bowl in Akron, Ohio.

University of Akron Photographs

Zippy with Softball Coach

One black and white photographic print of the Zippy the kangaroo mascot posing with a University of Akron female softball coach, probably at a softball game or practice in Akron, Ohio.

University of Akron Photographs

Zippy with Cheerleaders at Rubber Bowl

One black and white photographic print of Zippy the kangaroo mascot standing on the field at the Rubber Bowl in Akron, Ohio, probably during a timeout at a University of Akron football game. Two female University of Akron...
Sharepoint

● What is it?
  ○ Microsoft Office web-based collaborative platform that offers document management and storage space
  ○ Archival Services began transition from institutional servers to Sharepoint in 2019, and is still transitioning

● Does it work?
  ○ Advantages (remote access, cloud storage complies with NDSA and DPCMM models, access can be determined per member/guest, notify admins when files are changed/edited/deleted)
  ○ Challenges (does not accept all file formats, issues uploading, long loading times, viewing large files/folders, naming convention complications, metadata, checksums)

● What is its purpose?
  ○ Not recommended for storage of digital collections
  ○ Work better as collaborative platform for short term file sharing and documentation
### Library Archives

#### Archives

<table>
<thead>
<tr>
<th>Name</th>
<th>Modified</th>
<th>Modified By</th>
<th>Format</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 01</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>384303</td>
</tr>
<tr>
<td>Box 02</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>384522</td>
</tr>
<tr>
<td>Box 03</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>384348</td>
</tr>
<tr>
<td>Box 04</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>384184</td>
</tr>
<tr>
<td>Box 05</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>384458</td>
</tr>
<tr>
<td>Box 06</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>384535</td>
</tr>
<tr>
<td>Box 07</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>384328</td>
</tr>
<tr>
<td>Box 08</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>384603</td>
</tr>
<tr>
<td>Box 09</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>383974</td>
</tr>
<tr>
<td>Box 10</td>
<td>February 3</td>
<td>Orcutt,Zoe</td>
<td></td>
<td>383976</td>
</tr>
</tbody>
</table>

**Sharepoint screenshot**
Where we are now.
Know your content and your challenges to tailor your digital preservation policy to your library.
Resources

- **Frameworks for assessment**
  - NDSA Levels of Digital Preservation
  - Digital Preservation Capability Maturity Model
  - DPC Rapid Assessment Model
- Open Archival Information System
- Sustainability Health Check Tool for Digital Content Projects
- TheMaintainers.org - if you’re interested read this first!
- Three Legged Stool of Digital Preservation
- Klindt, M. PDF/A considered harmful for digital preservation, iPres 2017
- PDF Universal Accessibility (PDF/UA)
- UA Archives CONTENTdm Repository
- Using Sharepoint as a collaboration tool
Thank you!

Sidney Gao
gaosy@ucmail.uc.edu

James Van Mil
vanmiljf@ucmail.uc.edu

Brittany Hayes
bhayes@uakron.edu

Zoe Orcutt
zeo1@uakron.edu

https://bit.ly/2N1hQS2