Managing Your Digital Collection

MAGGIE DOWNING
MANAGER OF DIGITAL IMAGING
CONSERVATION CENTER FOR ART AND HISTORIC ARTIFACTS
Introduction

Maggie Downing, Manager of Digital Imaging
Conservation Center for Art and Historic Artifacts
mdowning@ccaha.org
What We’ll Discuss

- Challenges in managing digital collections
- Key terms
- Characteristics of digital records
- Basic activities in digital preservation
- Digital preservation self-assessment
Challenges
Dependence on Technology
Volume
Error or Attack
Feels Overwhelming
Key Terms

- Archive / Digital Record
- Digital Preservation
- Digitized / Born-Digital
- File Format
- Media
- Metadata
- Fixity / Checksum
Archive

Materials created or received by a person, family, or organization, public or private, in the conduct of their affairs that are preserved because of the **enduring value** contained in the information they contain or as **evidence** of the functions and responsibilities of their creator.

Digital Record

Data or information that has been captured and fixed for storage and manipulation in an **automated system** and that **requires the use of the system to render it intelligible by a person**.

Credit: Society of American Archivists
Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.

-American Library Association
Digitized

- Born-analog material
- Scanned or photographed from a physical format
  - Example: TIFF file from a negative, document, or book

Born-Digital

- Files created with software
  - Example: Word documents, InDesign layouts, databases, email, websites
File Format

- Conventions for encoding data into human-readable form
- Can be proprietary or open-source
- Examples: TIFF, JPEG, DOC, MP4, WAV, PDF
Media

- Where the file is stored
  - Hard Drive
  - Server
  - The Cloud
  - Digital Repository
  - Portable / Removable Media
Metadata

Information that describes, explains, locates, or makes it easier to retrieve, use, or manage the information resource

Some metadata mirrors what is created for physical records
- Descriptive – Dublin Core, MARC
- Administrative – Copyright and access restrictions

Some is specifically for digital records
- Technical – Scanner/camera, date created, pixel dimensions, etc.
- Structural – File’s relation to other files
- Preservation – Checksums, history of data corruption or recovery
Fixity / Checksum

- Fixity is the assurance that a digital file has remained unchanged
- Done by creating a checksum, or “digital fingerprint”
- Digital Preservation Coalition site on Fixity and Checksums
Characteristics of a Digital Record

- Authenticity
- Reliability
- Integrity
- Usability

Digital preservation is the ongoing maintenance of all characteristics over time.

Credit: Society of American Archivists
Authenticity

Can it be proven that the digital record is what it attests to be?

Authenticity can be established by adding metadata

- Administrative metadata
- Technical metadata
- Descriptive metadata
- Preservation metadata
Reliability

Is the digital record complete and accurate?

Reliability can be established through structural metadata
- Is the digital record part of a larger group? Does it represent one page in a group of letters?
Integrity

Is the digital record complete and unaltered over time or in transit?

Integrity can be established by:

- Computing a checksum
- Using a system that assigns a unique ID to all records, avoiding duplication
- Storing final files as “read-only”
Usability

Is the digital record accessible?

Usability can be established by:

- Using a consistent file and folder naming schema
- Creating indexes and inventories of the digital records
- Employing a system for search and retrieval (that is not just full-text search in a server, could use a CMS, DAMS or Excel sheet)
Digital Preservation Activities

◆ **Identify** digital content that you have
◆ **Select** content that warrants preservation
◆ **Process** the selected content by arranging, describing, and preparing it for storage
◆ **Store** selected content
◆ **Maintain** selected content over the long term through monitoring, migration, and recovery

Adapted from Library of Congress and Digital POWRR
Identify

◆ Goal: Take stock of the digital materials that are in your care.
◆ Strategy: Talk with staff, interns, volunteers, supervisors, IT as needed
◆ Tools: Excel, Access, Google Sheets, or other database tools
◆ Resulting Document: Digital asset register
Digital Asset Register: Information to Gather

- Name of collection / content
- Person / department responsible for maintaining the collection
- Size of collection
- Location of files
- Backup policy
- File formats
- Retention policy
- Ownership, rights, and data protection issues
- Associated risks
- Estimated value of content
Select

- **Goal:** Select which digital materials require “long-term preservation”
- **Strategy:** Talk with staff, interns, volunteers, supervisors, IT as needed
- **Tools:** Staff knowledge, mission statement, and other guiding documents
- **Resulting Document:** Digital Preservation Selection Policy
Selection Criteria

Selection criteria should:

▪ Support your mission

▪ Reflect criteria for preserving physical material

▪ Inform the creation of new digital content through digitization and collecting born-digital content

▪ Prioritize file formats that are widely adopted and not platform-specific
A Note on File Formats

- Widely adopted
  - Common file types that lots of people use are more likely to stick around

- Platform-independent
  - It can be opened in multiple programs

Library of Congress Recommended Formats Statement
https://www.loc.gov/preservation/resources/rfs/TOC.html

Smithsonian Recommended Preservation Formats for Electronic Records
https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records
Process

- Goal: Establish the characteristics of digital records
- Strategy: Begin to process digital materials before accession
- Tools: BitCurator, Archivematica, and others
  - POWRR Tool Grid: https://digitalpowrr.niu.edu/digital-preservation-101/tool-grid/
- Resulting Documents: Metadata procedures, data dictionary, donor agreements, access policies
Process: Steps

- Gather contextual information
- Perform a conservation assessment
- Identify access restrictions
- Arrange the records
- Describe the records
- Create access tools
Process: Gather Contextual Information

- How were materials created? Why?
- How were they previously managed?
- What is their current context?
- What hardware and software dependencies are there?
Process: Conservation Assessment

- Virus scan
- Identify and validate file formats
- Generate a checksum
- Identify preservation issues
Process: Identify Access Restrictions

- Identify Personally Identifiable Information (PII)
- Identify copyright status if possible
- Identify embargo restrictions
- Identify culturally sensitive information
Process: Arrange Materials

- Determine if digital records have an original order, and keep a record of this
- Identify relationships between groups of materials
- Rearrange files into series
Process: Describe Materials

- Does not need to be item-level
- Create descriptive, administrative, structural, and preservation metadata
Process: Metadata

How is metadata associated with a file?

◦ It can accompany the file as an Excel spreadsheet
◦ It can be embedded into the file itself through Adobe Bridge or other tools
◦ It can be input through a content management system such as TMS, PastPerfect, or ContentDM
Process: Create Access Tools

- Finding aids
- Online catalog access
- Online image access
Store

Goal: Establish or maintain authenticity and integrity of digital records

Strategy: Regularly perform fixity checks and assess security of storage system

Tools: Local Servers, cloud storage, digital preservation systems, digital repositories

Resulting Document: Storage and backup procedures
Store

- Perform a fixity check or generate a checksum if possible
- Store processed digital records in a “read-only” directory on your server
- Practice 3-2-1 backup procedure
  - 3 copies, 2 media types, and at least 1 copy maintained offsite
- Media types include:
  - Network servers
  - Cloud storage
  - Digital repositories
  - Removable media
Manage

- Goal: Ensure ongoing access to the digital records over time
- Strategy: Monitor files and migrate as needed. Monitor field of digital preservation
- Tools: Fixity
- Relevant Document: Digital Preservation Policy
Manage: Monitor Files

- Monitor files
  - Continue to perform fixity checks

- Perform preservation audits
  - Ensure that processing actions are meeting stated obligations

- Maintain technical infrastructure:
  - Performing maintenance on hardware, software, facilities, supplies, and technical components used for storage and access, as needed

- Migrate files
  - Convert data to latest file formats or relocate to new storage media as required
  - Consider a migration schedule every 5-10 years
Manage: Monitor Field

- Monitor the field of digital preservation
  - Digital Preservation Coalition – dpconline.org
  - American Library Association DigiPres listserv – https://lists.ala.org/sympa/info/digipres
  - Sustainable Heritage Network – https://sustainableheritagenetwork.org/
  - Lyrasis courses – https://www.lyrasis.org/services/Pages/Classes.aspx
  - POWRR – https://digitalpowrr.niu.edu/
National Digital Stewardship Alliance: Levels of Digital Preservation

◆ Tiered set of recommendations on how to build or enhance digital preservation activities
◆ A lightweight tool for self-assessment and to encourage organizations to think about digital preservation goals
◆ Categories include Storage, Integrity, Control, Metadata, and Content
<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>
</table>
| **Storage**     | Have two complete copies in separate locations  
Document all storage media where content is stored  
Put content into stable storage | Have three complete copies with at least one copy in a separate geographic location  
Document storage and storage media indicating the resources and dependencies they require to function | Have at least one copy in a geographic location with a different disaster threat than the other copies  
Have at least one copy on a different storage media type  
Track the obsolescence of storage and media | Have at least three copies in geographic locations, each with a different disaster threat  
Maximize storage diversification to avoid single points of failure  
Have a plan and execute actions to address obsolescence of storage hardware, software, and media |
| **Integrity**   | Verify integrity information if it has been provided with the content  
Generate integrity information if not provided with the content  
Virus check all content; isolate content for quarantine as needed | Verify integrity information when moving or copying content  
Use write-blockers when working with original media  
Back up integrity information and store copy in a separate location from the content | Verify integrity information of content at fixed intervals  
Document integrity information verification processes and outcomes  
Perform audit of integrity information on demand | Verify integrity information in response to specific events or activities  
Replace or repair corrupted content as necessary |
<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>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></td>
<td>Implement metadata standards chosen</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>
</tbody>
</table>
Conclusion

- There are many unique challenges in managing digital collections.
- When we understand the characteristics of a digital record, we can take actions to preserve them, and reduce the risks of these challenges in our collections.
- As we work toward our goals, it is helpful to periodically assess where we are and what we might take as next steps.
Questions?

THANK YOU!

EMAIL: MDOWNING@CCAHA.ORG