

Introduction to Digital Preservation

Danielle Mericle

dmericle@uoregon.edu

University of Oregon

What is Digital Preservation?

“...the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability, and accessibility of content over the very long term.”

—JISC, Digital Preservation Coalition, 2009

“Digital preservation combines policies, strategies and actions that ensure access to information in digital formats over time” —ALA working definition

Why Digital Preservation?

- ▣ Guaranteed access to digital contents in perpetuity
- ▣ Return on investment
- ▣ Management and discoverability of digital assets across organization
- ▣ Rights management
- ▣ Responsible stewardship

Goals of Digital Preservation

- Creation of digital objects that are readable, authentic, and understandable over time
- Vetted digital preservation policy
- (Ideally) a technical solution that can help you manage your digital resources over time

Where do we start?

- Criteria of preservation “readiness” as defined by National Digital Stewardship Alliance ([NDSA](#))
 - Storage & Geography
 - File Fixity & Data Integrity
 - Information Security
 - Metadata
 - File formats

Storage & Geography

- Maintain at least two, ideally three, copies of your digital files
- Utilize different storage mechanisms (don't rely on one form of technology)
- Separate your copies geographically (think natural disaster)
- Separate your master copies from your access copies
- When using localized media (disks, external drives), use common-sense handling- minimize dust, jarring, temperature fluctuations, magnets, UV,
- Migrate assets regularly (*external drives have a shelf-life of 3-5 years)

File Fixity & Data Integrity

- Generate checksums, or “digital snapshots” of your files, as you digitize your holdings
(<https://www.avpreserve.com/tools/fixity/>)
- Virus check your content (especially before moving to long-term storage)

Information Security

- ▣ Identify who within your organization has the right to view, modify, delete content
- ▣ Document who has accessed content and when

Metadata

- ▣ Create an inventory of your digital assets
- ▣ Capture as much administrative (project) information as possible
- ▣ Capture technical metadata regarding capture or origin of files
- ▣ Integrate or link to descriptive information

Formats

- Know your formats! Document them.
- Select recommended preservation formats as established by national and international consortiums
(<http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>)
- Monitor your formats for obsolescence issues

Baseline preservation requirements

- Standardize file formats & naming conventions
- Maintenance of “master” images separate from “access” images
- Plan for refreshing and/or migrating data
- Metadata capture and management
 - Technical- device; color management info; date; operator; format; inhibitors
 - Descriptive- any bibliographic info; existing or new data
 - Administrative- project info; provenance; history
- Redundant storage (x3)
 - Geographic separation and media variety

A digital preservation policy

- Clearly articulates roles and responsibilities amongst organization
 - Who maintains/migrates data (back-end)
 - Ensures access to data (front-end)
 - Monitors financial well-being of preservation system
- Defines scope and length of preservation
- Defines ultimate accountability

Policy questions

- Who coordinates & approves preservation policies? Copyright compliance?
- Who creates and manages content?
- What file formats will you support?
- What content will be retained and for how long?
- Who can submit to the repository?
- Who can download or view content?
- What metadata standards are used?
- Who owns the content?
- How is copyright and IP controlled?

Preservation Repository Basics

- ▣ Ingest
- ▣ Processing
- ▣ Access
- ▣ Storage
- ▣ Maintenance
- ▣ Other

Preservation repository functions

Digital POWRR Tool Evaluation Grid	Ingest					Processing					Access		Storage				Maintenance			Other			
	Copy	Fixity Check	Virus Scan	File Dedupe	Auto Unique ID	Auto Metadata Creation	Auto Metadata Harvest	Manual Metadata	Rights Management	Package Metadata	Auto SIP Creation	Public Interface	Auto DIP Creation	Auto AIP Creation	Reliable, Long-Term Bit Preservation	Redundancy	Geographically Dispersed Data Storage Model	Exit Strategy	Migration	Monitoring	Auto Recovery	Open Source	Clear Documentation

Repository packages

- Submission Information Package (SIP): metadata and content data files to be ingested in repository
- Archival Information Package (AIP): ingested SIP that may have been normalized/localized for long-term storage
- Dissemination Information Package (DIP): Export package of AIP, including “last, best” version of any file

Vendor solutions

- ❑ Not all solutions provision for all stages of digital preservation (ingest to storage to access)
- ❑ Open-source doesn't equate with "free"
- ❑ Evaluate requirements analysis across your institution
 - ❑ Scope
 - ❑ Technical and staff capacity
 - ❑ Funds
 - ❑ Collection analysis

Some popular vendor solutions

- ▣ Archivemática
- ▣ Preservica
- ▣ Archives Direct
- ▣ Internet Archive
- ▣ Many others...

Preservation plan

- ✓ Develop relevant policies
- ✓ Determine selection- content types, scope, etc
- ✓ Define roles & responsibilities
- ✓ Secure institutional support/funding
- ✓ Benchmark
 - ✓ Format types and settings
 - ✓ Metadata types, minimum requirements
- ✓ Migration, refreshment strategies
- ✓ Storage, including back-up / redundancy
- ✓ Copyright compliance, access and permissions, both for ingest and download
- ✓ Generate requirements document
- ✓ Determine your solution for integrated technical architecture

Final thoughts

- ▣ Start small, scale up
- ▣ Aim to achieve best practices, at minimum
- ▣ Advocate your case
- ▣ Consider partnering with other institutions – preservation can be expensive
- ▣ 3-3-3 Digital Preservation Action Plan

Resources

- Sustainable Heritage Network:
<http://sustainableheritagenetwork.org/>
- POWRR, Preserving (Digital) Objects with Restricted Resources:
<http://digitalpowrr.niu.edu/>
- National Digital Stewardship Alliance (NDSA):
<http://ndsa.org/>
- Library of Congress Digital Preservation:
<http://www.digitalpreservation.gov/>
- 3-3-3 Digital Preservation Action plan
<http://sustainableheritagenetwork.org/digital-heritage/3-3-3-digital-preservation-action-plan>