Welcome to CARLI Digital Preservation Webinar Series: Store & Protect

We will begin the presentation at 10:00 am.

If you have not already run the Audio Setup Wizard, you can do so now. In this meeting room, go to the top left menu item called Meeting -> Audio Setup Wizard.

If you would like to ask a question during the presentation, please type your question into the Chat box on the left side of the screen.
Digital Preservation: Store & Protect

Considerations for long-term storage and protection of digital objects
Store, part 1

Laurie Sauer
Knox College
Six Steps of Digital Preservation

• **IDENTIFY** the types of digital content you have.

• **SELECT** what portion of your digital content will be preserved.

• **STORE** your selected content for the long term.

• **PROTECT** your content from everyday threats and emergency contingencies.

• **MANAGE** and implement requirements for long term management.

• **PROVIDE** access to digital content over time.
Digital objects may be stored, but are they being preserved?
Multiple copies in at least 2 locations
Common (or normalized) file formats
Identification and description about each object: Metadata
Controlled and known storage of content
How many copies are enough for you?

Minimum: 2 copies in two locations
Optimum: 6 copies

Storage factors:
- Video files are too large to store 6 copies
- Possible legal restrictions
- Types of media used for storing the content
Follow recommendations set by leading organizations.

- NARA’s *Technical Guidelines for Digitizing Archival Materials for Electronic Access* – TIFF format is the “‘De facto’ raster image format used for master files.”
  http://www.archives.gov/preservation/technical/guidelines.html

- *Sustainability of Digital Formats Planning for Library of Congress Collections* -- The MP3 sound file format is “Generally used for final-state, end-user delivery.” And, “General preference for preservation-oriented recorded sound is WAVE_LCPM. For compressed sound, MP3 is acceptable, especially at data rates of 128 Kb/s (mono) or 256 Kb/s (stereo) or higher.”
  http://www.digitalpreservation.gov/formats/index.shtml
What is stored?

- file
  + metadata
- digital object
- **Environment** – information required to access, render and use the object

- **Rights management** – information that describes current and future use restrictions

- **Provenance** - descriptions of actions that have been taken to preserve the object over time, including actions that alter the content; includes information that validates object's authenticity, e.g., fixity checks

From *Preservation Metadata* (2nd ed.)
http://dx.doi.org/10.7207/twr13-03
Preservation Metadata
Content (what), Fixity (unchanged), Provenance (life story), Reference (this thing), Context (relationships)

Administrative (manage)
Structural (understand, use)
Descriptive (find, use)
Title: [Colombian Liberation]
Description: Various Colombian martyrs are portrayed in this mural. The background of this mural is the colors of the Colombian flag; text in the background is indistinct.
Photographer: Houser, Henry P.
Identifier: The Henry Houser Collection (Part 3 Folder 12:1)
Rights: U.S. and international copyright laws protect this digital image. Commercial use or distribution of the image is not permitted without prior permission of the copyright holder. For permission to use the digital image, please contact Knox College Special Collections & Archives at archives@knox.edu - http://www.knox.edu/library/special-collections-and-archives.html
Collection: Muralism in Revolutionary Nicaragua-The Henry Houser Collection (Knox College)
Fixity checking allows you to know if a file has changed over time.

Store

Preservation metadata

One-Way Encryption

b43efderwkl3jh7834

One-Way Encryption

845kjsnlkdrkjhndgiu5
Minutes from trustees’ meetings from the 1980s were migrated from MS Word to PDF/A.
Store, part 2

Mary Z. Rose
Southern Illinois University Edwardsville
1. Make multiple copies of your stuff
   How many copies? 2 is good, 6 is best

2. Keep the copies in different geographic locations

Photo by djaquay on Flickr http://www.fotopedia.com/items/flickr-1526855132
Store

Storage media options

Offline
Near-line
Online

Photo by Anonymous Account on Flickr http://www.fotopedia.com/items/flickr-366070621
Good
1. Easy to keep up to date
2. Multiple copies are easy to achieve
3. Easy to access (or is this a Not So Good thing?)

Not So Good
1. More security issues from vandalism
2. Higher costs long-term
Storage partners

Online storage options

Hosted services (cloud)

- Meta Archive
- Preservica
- Chronopolis
- Academic Preservation Trust
So you want to host your own digital asset preservation management repository?
Resources

General

Digital Preservation Coalition Technology Watch Reports
http://www.dpconline.org/publications/technology-watch-reports
Very readable reports by experts on issues of digital preservation

The Signal: Digital Preservation
http://blogs.loc.gov/digitalpreservation/
Library of Congress blog on technology and digital preservation topics

Reference Model for an Open Archival Information System
http://public.ccsds.org/publications/archive/650x0m2.pdf
More (a LOT more) information about digital preservation goals and standards
Preservation Metadata

*Preservation Metadata (2nd ed.)*, by Brian Lavoie and Richard Gartner
http://dx.doi.org/10.7207/twr13-03
A DPC Technology Watch Report; very informative and readable

**PREMIS**
http://www.loc.gov/standards/premis/
The Library of Congress maintains the PREMIS Data Dictionary standard for preservation metadata

**NC Preservation Metadata for Digital Objects**
An example from the State Library of North Carolina of PREMIS implementation.
Resources

Technical

About file formats:
• NARA’s Technical Guidelines for Digitizing Archival Materials for Electronic Access
  http://www.archives.gov/preservation/technical/guidelines.html
• Sustainability of Digital Formats Planning for Library of Congress Collections
  http://www.digitalpreservation.gov/formats/index.shtml

A comprehensive resource for data file formats and associated software.

Digital POWRR Tool Grid -- http://digitalpowrr.niu.edu/tool-grid/
Evaluations of digital preservation tools.

Two good explanations of fixity checking:
• Checksum video: http://www.youtube.com/watch?v=Emom_ncMqu0
• “Hashing Out Digital Trust” – a Blog post on The Signal about hash functions
Storage

Another DAM blog
http://anotherdamblog.com/
A vendor-neutral blog about digital asset management.

Review of Available Open Source DAM Software
(Naresh Sarwan August 10, 2013)
http://www.opensourcedigitalassetmanagement.org/reviews/available-open-source-dam/

Report on Digital Preservation and Cloud Services
(Minnesota Historical Society, April 1, 2013)
Storage Options

Near-line storage
  • Amazon Glacier http://aws.amazon.com/glacier/

Networked storage cooperatives
  • Meta-Archive http://www.metaarchive.org/
  • Chronopolis http://chronopolis.sdsc.edu/

Hosted services for online digital preservation storage
  • Preservica http://preservica.com/
  • DuraCloud http://www.duracloud.org/
Storage Options (cont’d)

Open source digital asset management (DAM) software for preservation

• Concerto http://concerto.sourceforge.net/
• DSpace http://www.dspace.org/
• Fedora Commons http://www.fedora-commons.org/
• Greenstone http://www.greenstone.org/
• DAITSS http://daitss.fcla.edu/
Identify - what digital content do you have?

Select - what portion of that content will be preserved?

Store - what issues are there for long term storage?

Protect - what steps are needed to protect your digital content?

Manage - what provisions are needed for long-term management?

Provide - what considerations are there for long-term access?
What are we protecting content from?

- Change and loss – accidental and intentional
- Obsolescence – as technology evolves
- Inappropriate access – e.g., confidential data
- Non-compliance – standards and requirements
- Disasters – emergencies of all kinds
Disasters – emergencies of all kinds
Dangers to digital objects

Non-compliance—standards and requirements
Non-compliance—standards and requirements
Inappropriate access – e.g., confidential data
Dangers to digital objects

- Digital content must be subject to security measures just like analog materials
Dangers to digital objects

Obsolescence – as technology evolves
Dangers to digital objects

Obsolescence – as technology evolves
Dangers to digital objects

Change and loss – accidental and intentional

These black Write Protect Tabs should be used when you do not want stored information on your disk to be changed or lost accidentally. Simply cover the Write Enable Notch by taking 1 black tab and wrapping it around the notch.
Dangers to digital objects

Change and loss – accidental and intentional

- Mary Shelly’s *Frankenstein*
- *Mona Lisa*
How digital things are the same

- Formal sense: same ones and zeroes
- Forensic sense: how the bits are physically encoded and inscribed on an object
Protect

Dangers to digital objects

Forensic tools

Hardware write-blocker

FTK Imager
Benn Joseph
Manuscript Librarian
Northwestern University Library
b-joseph@northwestern.edu
Protect, part 2

Adam Strohm
Newberry Library
**Know** where your content is located
- onsite vs. offsite, online vs. offline (via administrative metadata)

**Know** who can have access to it
- Which staff members? Which departments? (via permissions)

**Know** who accesses your secure information
- Staff, depositors, users? (via authentication)

**Know** about your users to improve service
- Web use, internal use, user activities, maintenance (via user data)
Proper planning should allow you to:

• **Prevent** undesirable outcomes
• **Predict** the most likely risks and threats
• **Detect** errors, problems, and damage
• **Respond** with appropriate measures
• **Repair** damage or possible loss

**CAUTION** **CAUTION** **CAUTION** **CAUTION**
Steps to protect your content

• Identify possible risks
• Define those risks (in nature and scope)
• Assess potential impact and possible damage
• Develop appropriate, feasible response plans
• Respond to risks and threats (implementation)
Engage in ongoing disaster planning

- Establish a committee and share information
- Develop and maintain documents
- Update plans regularly

Identify possible outcomes and prepare

- e.g., server failure, media damage, data loss
- Practice! (simulations and drills)
What needs to be available soonest?

- Identify core functions as part of planning
- Determine allowable downtime for each
- Consider steps to re-establish each function
- Develop relevant documents
- Make sure planning documents are accessible
dPlan
• Free online disaster planning tool from the NEDCC and MBLC

TRAC (Trusted Repository Audit and Checklist)
• Criteria and checklist from CRL

TDR (Trusted Digital Repository checklist)
• ISO standard based largely on TRAC

DRAMBORA (Digital Repository Audit Method Based On Risk Assessment)
• Interactive digital auditing tool from the Digital Curation Centre

DIGITAL POWRR (Preserving Digital Objects with Restricted Resources)
• Ongoing collaborative assessment of digital preservation planning for smaller institutions
Protect

Planning Components

- IT contingency
- Staff coverage
- Crisis communication
- Operational continuity
- Hacking/virus response
- Disaster recovery

Adapted from NIST Contingency Planning Guide for Federal Information Systems
Good practice should result in

• Practices in place to manage day-to-day protection (an implemented preservation policy)
• Disaster planning in place to prevent, predict, respond, and repair (preparation in the event of an emergency)
  http://www.gpo.gov/authentication

• Data Seal of Approval: http://www.datasealofapproval.org/

• MetaArchive Trusted Repository Audit (2010):

• Edinburgh Data Audit Implementation Project Final Report (2009):
  http://ie-repository.jisc.ac.uk/283/1/edinburghDAFfinalreport_version2.pdf


Additional Resources

- dPlan: www.dplan.org
- Drambora: www.repositoryaudit.eu
- TRAC & TDR: www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0
- Digital POWRR
digitalpowrr.niu.edu/
Thank you!

Benn Joseph  b-joseph@northwestern.edu
Mary Rose    mrose@siue.edu
Laurie Sauer  lsauer@knox.edu
Adam Strohm  strohman@newberry.org
Thank you for attending!

Please complete the online evaluation:

https://www.surveymonkey.com/s/digipres_storeprotect

The CARLI Digital Preservation Trainers will use your feedback when planning future webinars and events.