Library Preservation Today! Preserving Digital Collections: An Overview

AN ALCTS WEBINAR PRESENTED BY

PETER D. VERHEYEN
SYRACUSE UNIVERSITY LIBRARIES/
CONSERVATOR IN PRIVATE PRACTICE

MARCH 16, 2016

ALCTS
www.ala.org/alcts

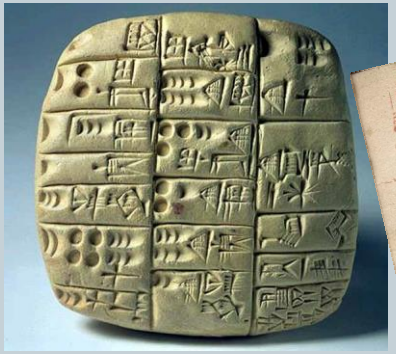
Peter D. Verheyen

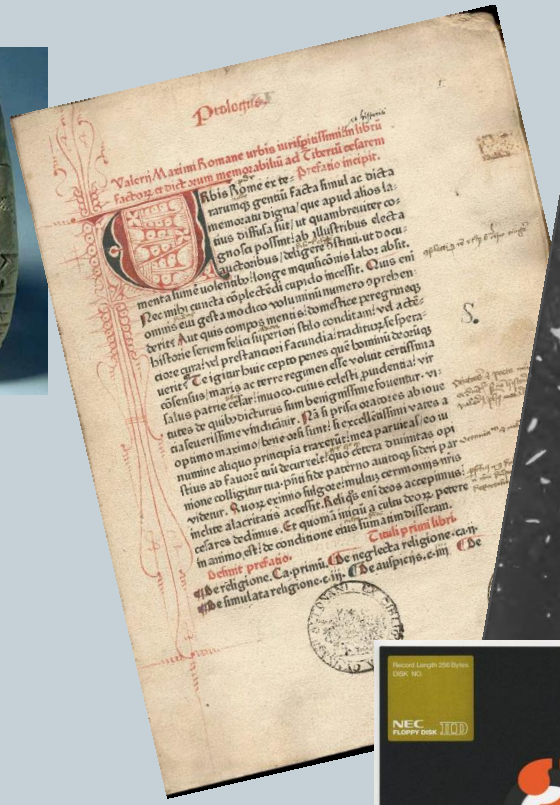
Peter D. Verheyen, Research and Emerging Issues Analyst at Syracuse University Library comes from a bookbinding and rare book conservation background. He was trained in hand bookbinding in Germany and worked as a conservator in private practice in Chicago before going to work as conservator at Yale and Cornell University libraries. It was there that he was first exposed to digitization in the early 1990s. In 1995 he established the conservation lab at the Syracuse University Libraries where he also served as Head of Preservation and Conservation until 2013. Verheyen was involved in digitization activities since coming to Syracuse, managing many of the digitization projects in the Special Collections Research Center. He completed both of Cornell's workshop series on digitization and digital preservation and has presented on various aspects of the digitization process including scanning, project management, the wider impact of digitization on public services and preservation. He would like to see digitization applied more systematically to the preservation of endangered collections as well as the preservation of the digital.

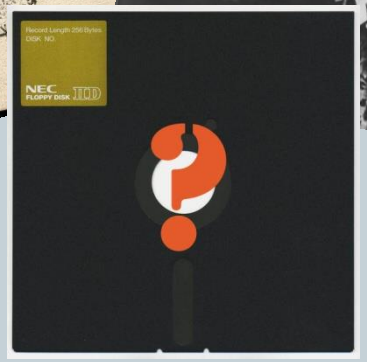
So, what have we learned in this series?

- Importance of environmental monitoring and control to prevent damage to collections
- Standards for library storage, and ways of achieving better conditions, including low- and no-cost improvements
- Storage Furniture (Shelving and related issues)
- The Storage Area
- Storing and Handling Books
- How Scrapbooks Differ from Books
- Storing and Handling Scrapbooks

Historical Context







Historical Context

- Rock
  - Cave paintings, stone carvings, clay tablets
- Paper
  - Papyrus, parchment/vellum (not paper, but), rag, wood
  - “Brittle book” problem
  - Standards developed to ensure longevity
- Microfilm
  - Early film quality spotty, but…
  - Standards developed to ensure quality (chemical & image)
- Despite problems, all remain readable

What digital preservation is not.

Digital preservation is not the digitization of analog objects, however the formats and standards used for digitization will significantly impact long-term preservation of the digital objects.

What digital preservation is

Digital preservation, is a part of the curation process and the management of digital assets over time to ensure their continued accessibility.

Other definitions at end of slides

Quiz 1

- Are you currently digitizing collections
  - No
  - Yes, still images
  - Yes, audio
  - Yes, video
  - A combination of these

A word with eight letters

- Metadata, data about data
  - Don't digitize without it
  - A critical piece before, during and after digitization
    - Descriptive: Describes the object for discovery…
    - Structural: Describes how complex objects relate to each other
    - Administrative: Enables the management of the object. Includes:
      - Technical: Info to render, interact with and use the digital object
      - Rights: Copyright…
      - Preservation: What do we need to know to preserve. Can include elements of all of the above.
  - More in other parts of presentation

PREMIS Data Dictionary. http://www.loc.gov/standards/premis/
Preservation Metadata (2nd edn.). Lavoie, Brian and Gartner, Richard (2013). http://dx.doi.org/10.7207/twr13-03

Types of digital objects

- Digitized analog
  - Text, images, audio, video, film on a variety of media
- Born-digitized
  - Computer files in different formats and on various media
    - Tape, disks, cd/dvd, servers
    - Programs and software
      - Proprietary
      - Open Source
      - Unknown – old media often not or erroneously labeled…
  - Websites, email, “social media,” …

What can we do to facilitate preservation?

- Adhere to standards and best practices when creating digital content by:
- Digitizing at highest appropriate and/or practical resolution
- Creating use-neutral masters in lossless, stable, non-proprietary formats: TIFF, JPEG2000, WAV (audio), AVI (video)…
- May not always have control with digital devices like recorders, phones, …
- If skills, budget, tools to complex/$$$, outsource to professional digitization vendors

What can we do to facilitate preservation?

- Write-protect master images to prevent altering, working off of derivative copies instead
- Use logical naming conventions and gather metadata as you work on collections
- Do not save/"preserve" on media such as CDs, DVDs, flash drives, other proprietary storage device
- Rights management is omnipresent

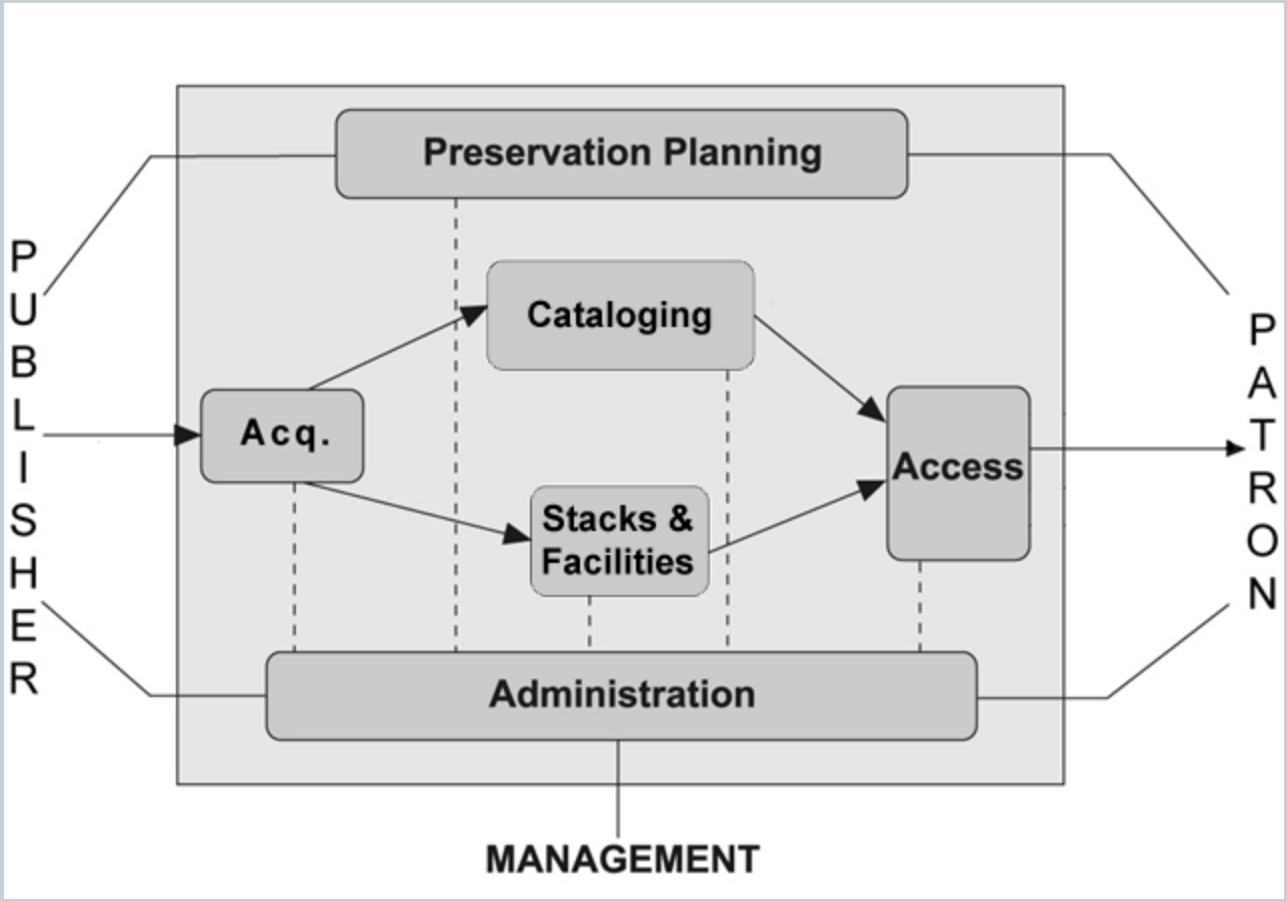
Digital Preservation Best Practices and Guidelines (in NC), see http://digitalpreservation.ncdcr.gov/

Let's preserve…

Quiz 2

- We are currently storing our digital objects on: Check boxes
  - Media
  - Hard drives
  - Servers / In the cloud
  - We aren't

Analog management chain



Based on OAIS model

Digital content is fragile

- Media and storage devices degrade and/or become obsolete quickly.
  - Replacement costs…
  - Appeals for old hardware/software…
- Software, file formats, and operating systems become obsolete.
  - New versions may not be backwards compatible
- Proprietary encoding schema disappear.
- Files are deleted…
  - Early working group on DP listserv archived and lost at Yale
- Web links 'break.'

Digital preservation strategies & tools

- Storage
  - Essential, but only a first step… Backups are not preservation!
- Refreshing
  - involves periodically moving a file from one physical storage medium to another to avoid the physical decay or the obsolescence of that medium
- Migration
  - involves periodically moving files from one file encoding format to another that is useable in the current computing environment.
- Emulation
  - mimics obsolete applications software to run in the current computing environment
  - Common with video/arcade games

Curation is…

- Integral part of managing collections whether analog or digital
- Critical for ensuring success of [digital] preservation
- Other digital preservation components include: preservation, archiving and storage
  - Curating the Analog, Curating the Digital
    - http://www.archivejournal.net/issue/3/archives-remixed/
  - Video from CLIR (Council on Library and Information Resources) at http://www.clir.org/initiatives-partnerships/data-curation

Planning for digital preservation

- Develop a plan for managing each file format in a digital archive.
  - May include specific strategies based upon data format risk assessments
    - How likely is support for software/hardware to disappear
  - Establish Digital Preservation policies
    - Involve bibliographers, catalogers, and faculty to determine selection criteria for assets to be preserved
    - Incorporate DP into existing collection development policies and workflows
      - Directory structure and naming conventions
    - Policies must address Intellectual Property rights and access.
    - Document best practices
    - Develop preservation metadata guidelines

Planning for digital preservation

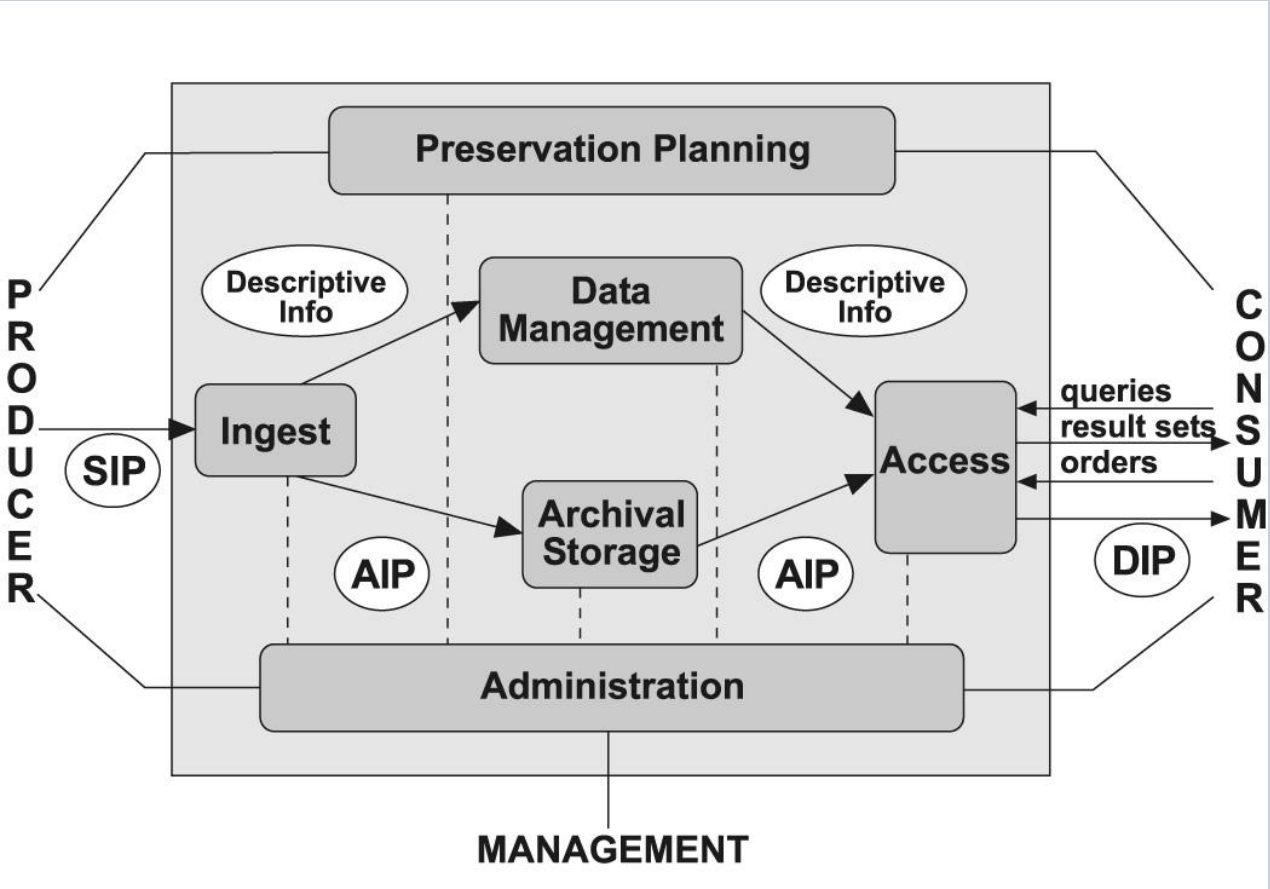
- Subject to regular review/revision
- Should be a collaborative effort between departments or among organizations
- Requires resources and organizational commitment
  - Raise awareness of DP among library staff and university community
  - Encourage/require good habits and standardization
    - File naming conventions
    - Directory structure

Sustainability

Trusted Digital Repository (TDR)

- Accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users;
- Have an organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility;
- Demonstrate fiscal responsibility and sustainability;
- Design its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it;
- Establish methodologies for system evaluation that meet community expectations of trustworthiness;
- Be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly;
- Have policies, practices, and performance that can be audited and measured; and
- Understand and act on requirements keeping in mind:
  - The scope of collections; preservation and lifecycle management; the wide range of stakeholders; ownership of material and other legal issues; and cost implications.

OAIS model ties all together



OAIS is…

- A “basic” framework for understanding the concepts of long-term digital preservation
- Not proscriptive and guides processes to allow institutions to implement digital preservation in ways that work for their circumstances in sustainable ways
- Wiki provides a clear and full description at:
  - http://en.wikipedia.org/wiki/Open_Archival_Information_System

Examples of DP initiatives and tools

- Co-operative Projects
  - LOCKSS & CLOCKSS
  - Portico
  - APTrust (Academic Preservation Trust) and DPN (Digital Preservation Network)
- Hosted Options
  - Digital Archive @ OCLC
  - DuraSpace (Fedora Commons and D-Space)
  - DSpaceDirect (hosted Digital Preservation from D-Space)

These DP initiatives…

- Preserve published scholarly outputs
  - Journal articles
  - eBooks
  - Locally digitized content
- Use different conceptual models
- Are largely cooperative and scalable
  - Distribute costs and responsibilities across several partners
  - Libraries , consortia, publishers

Examples of DP initiatives and tools

- LOCKSS/ CLOCKSS (Lots of Copies Keep Stuff Safe): "Grass roots" international community initiatives that provides libraries with digital preservation tools and support so that they can easily and inexpensively collect and preserve their own copies of authorized e-content.
  - Software migrates content forward in time; and the bits and bytes are continually audited and repaired.
  - Files are preserved as originally published (PDF, HTML, …) and migrated on access.
  - Can be used to create private/small consortial networks
  - CLOCKSS works more with publishers

Examples of DP initiatives and tools

- Portico: Preserves scholarly literature published in electronic form. Working closely with publishers,
  - Portico creates a dark archive from source files that have been converted to a standard format.
  - Archive is migrated forward en-masse as formats change. A standardized format
  - Similar mission as CLOCKSS.

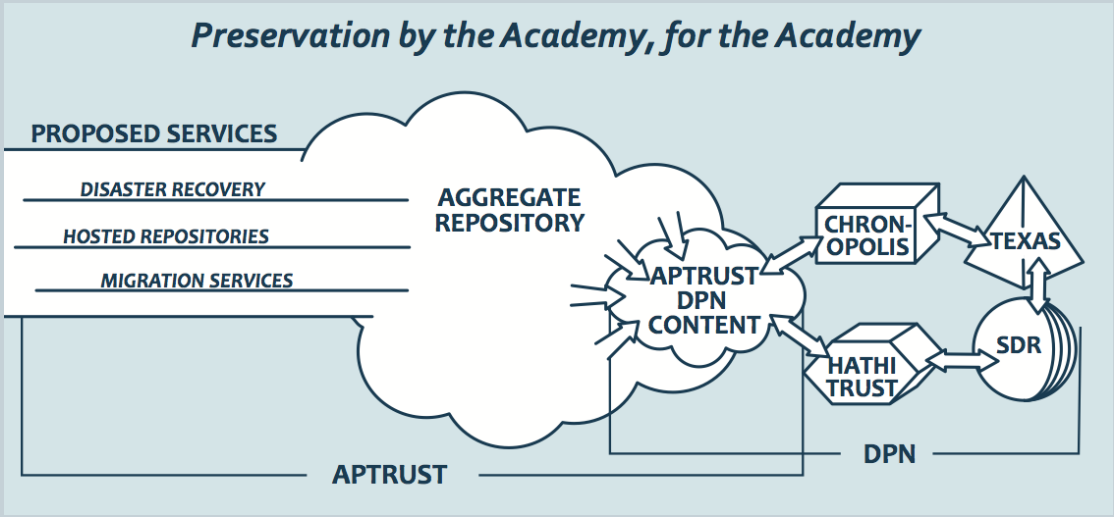
The e-journal challenge

- LOCKSS'/CLOCKSS' and Portico's "collection development missions have moved much closer together and there is no a great deal of overlap between them. This regrettably does not increase the number of titles being preserved…
  - We end our report by recalling our response to Question 10, in which we noted that LOCKSS and Portico combine to preserve roughly 26% of Cornell e-journals with standard identifiers and roughly 13% of all Cornell's e-journals. This overall lack of publisher participation in either of the leading e-journal preservation programs offers the two CULs an opportunity to use their individual or combined influence with publishers, to whom they pay substantial licensing fees, to improve the state of e-journal preservation as a whole.

  - Final Report of the 2CUL LOCKSS Assessment Team
    Cornell University Library & Columbia University Library
  - https://www.2cul.org/files/2CULLOCKSSFinalReport.pdf

DP for locally created collections

- AP Trust and DPN
  - "New " cooperative initiative for [largely] locally created collections
  - Based on replicating nodes running on different platforms



From
http://aptrust.org/2013/01/24/update-visits-to-aptrust-institutions-completed/

Academic Preservation Trust (APTrust)

- APTrust is consortium of academic libraries that is committed to the creation and management of a multi-institution collaborative preservation repository for digital content. Begun at the University of Virginia, the project began with 12 partners from private and public universities that are collaborating with Fedora, DuraCloud, and Cloudsync on development of the software.
  - Will be preservation AND access tool
  - Will include forward migration of "standard" file types

Digital Preservation Network (DPN)

- Digital Preservation Network is conceptually aligned with APTrust, DPN serves as a dark preservation archive constituting several federated and replicating nodes. To help ensure preservation, each of these nodes is runs on a different platform, APTrust, Chronopolis, HathiTrust, Stanford Digital Repository (SDR), and University of Texas Digital Repository.
  - Different platforms and replication ensure that there can be no single point of failure, and that if one node crashes, it can be "brightened" (recreated/refreshed) from the others.
  - Format migration is not planned as a feature of DPN.

Other DP tools

- OCLC Digital Archive provides foundation for preservation of digital collections:
  - Secure managed storage
  - Automated monitoring and workflows
  - Works seamlessly with CONTENTdm
  - Hosted, i.e. "ideal" for those without infrastructure
- DSpaceDirect is a low-cost, turnkey hosted repository service. DSpaceDirect can be used to preserve and provide access to academic faculty and student papers, projects, and research.
- Other IR-type cloud-based or locally hosted solutions such as DuraSpace and DuraCloud

So, why preserve?

- Digital preservation as insurance
- Preserved content is released when a “trigger event” occurs, e.g. publisher ceases operations or its delivery platform fails.
  - Release of a title can via LOCKSS/CLOCKSS or Portico can be temporary or long term depending on circumstance. APTrust and DPN will function similarly. So, when does preserved content become available?
  - HathiTrust (a slightly different beast) content is available as allowed by copyright.
    - Access may be made available locally for items physically held the institution if item is brittle/damaged beyond use.
    - Access may be made available locally for items physically held the institution for patrons with special needs (screen reader…)

So, … what can you do – "easy" steps

- Learn about what a digital archivist does in this OCLC Research blog post
  - http://hangingtogether.org/?p=3901
- Know what you have as content and media types
- Organize and maintain directory structure
- Limit "archival" formats to stable/established types
- Migrate forward (or backward) as formats change
  - Lowest common denominator
- Make use of metadata
- Keep control (intellectual & physical) of content
  - Keep your head out of "the cloud" except for access, for now
- Keep and maintain obsolete hardware as needed… you just might
- Find partners to share resources and costs

Thank you and questions

- Slides that follow are an extensive list of resources, including articles, tutorials , videos, and webinars.
- On behalf of my colleagues Karen Kiorpes, Julie Mosbo, and Donia Conn, thank you for attending this ALA/ALCTS series on preservation topics/

  - Peter D. Verheyen
    Syracuse University Libraries
    (315) 443-9756
    pdverhey@syr.edu
  - Website

Digitization resources

- *Moving Theory into Practice*, Cornell University Library.
  - https://www.library.cornell.edu/preservation/tutorial/
- Handbook for Digital Projects: A Management Tool for Preservation and Access, Northeast Document Conservation Center
  - http://nedcc.org/free-resources/digital-preservation
- Federal Agencies Digitization Guidelines Initiative (FADGI) recently
  - http://www.digitizationguidelines.gov/still-image/

Digital preservation resources

- National Digital Information Infrastructure Preservation Program (NDIIP)
  - "NDIIPP is based on an understanding that digital stewardship on a national scale depends on public and private communities working together. The program has engaged hundreds of partner organizations across the United States and around the world to preserve at-risk digital collections and build a distributed digital preservation infrastructure. This work is carried out through a variety of initiatives. A major current initiative is the National Digital Stewardship Alliance, which works to bring a broad array of organizations, both public and private, into partnership with the Library to support digital preservation.
  - Addresses "personal digital archiving"
  - http://www.digitalpreservation.gov/

Digital preservation resources

- Definitions of Digital Preservation prepared by the ALCTS Preservation and Reformatting Section, Working Group on Defining Digital Preservation: http://www.ala.org/alcts/resources/preserv/defdigpres0408
- Digital preservation at the Library of Congress: http://www.digitalpreservation.gov/
  - See especially *The Signal*, *http://blogs.loc.gov/digitalpreservation/*
- Preserving Objects With Restricted Resources: for libraries with smaller amounts of data and/or fewer resources. http://digitalpowrr.niu.edu/
- Community Owned digital Preservation Tool Registry (COPTR) http://coptr.digipres.org
- Digital Preservation Best Practices and Guidelines, see http://digitalpreservation.ncdcr.gov/
- Preserving Digital Information: Final Report and Recommendations published by the commission of the Commission on Preservation and Access: http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file2456.html
- Cornell's excellent online tutorial on digital preservation: http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html
- For more on attributes of Trusted Digital Repositories, see http://www.oclc.org/research/activities/trustedrep.html
- Preserving Digital Collections, UK National Archives, see http://www.nationalarchives.gov.uk/archives-sector/digital-collections.htm
- Preservation Metadata and OAIS, see http://www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf

Digital preservation resources

- Preservation Health Check: Monitoring Threats to Digital Repository Content, see http://www.oclc.org/research/publications/library/2014/oclcresearch-preservation-health-check-2014.html
- For examples of collaborative digital preservation efforts, see http://www.hathitrust.org/ and http://www.lockss.org
- For examples of digital preservation tools, see http://fedorarepository.org/about and http://www.dspace.org/
- Meghan Bergin. *Summary of Survey Results on Digital* Preservation Practices at 148 Institutions. http://works.bepress.com/meghan_banach/7/

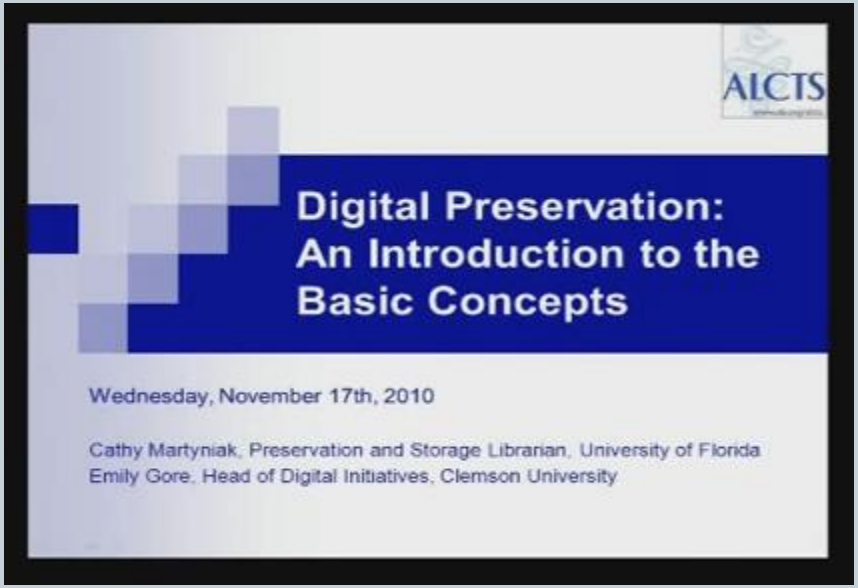
- Popular articles
  - The New Age: Leaving Behind Everything, Or Nothing At All. http://www.npr.org/blogs/alltechconsidered/2014/04/09/300614977/the-new-age-leaving-behind-everything-or-nothing-at-all
  - Jefferson Bailey Blog, http://www.jeffersonbailey.com/
  - The Library of Congress Wants to Destroy Your Old CDs (For Science), http://www.theatlantic.com/technology/archive/2014/05/the-library-of-congress-wants-to-destroy-your-old-cds-for-science/370804/
  - Tips on Archiving Family History, http://www.nytimes.com/2013/05/29/booming/tips-on-archiving-family-history-part-1.htm l www.nytimes.com/2013/06/05/booming/tips-on-preserving-family-films-and-photos.html http://www.nytimes.com/2013/06/12/booming/tips-on-archiving-family-history-part-3.html

Resources

- Digital Preservation In a Box
  - The latest project from the National Digital Stewardship Alliance, Outreach Working Group.
  - http://blogs.loc.gov/digitalpreservation/2012/07/digital-preservation-in-a-box-have-a-look-inside/ and http://dpoutreach.net/

Multimedia resources

Digital Preservation: An Introduction to the Basic Concepts



http://youtu.be/RqacRC51CRI

Preserving Your Personal Digital Memories

http://youtu.be/pDM5fLRWE4s

Format obsolescence

By 2036, data loss has become a thing of the past. All digital media is instantly uploaded to the internet and permanently stored in the cloud, safely backed-up on servers scattered around the world. Only a handful of small businesses in the world have the expertise to recover data from pre-cloud devices. On a hot summer day, a young man named Kai visits Digital Antiquities, a store in eastern Pennsylvania specializing in data recovery and sales of vintage electronics. He shows Cat, the stores only employee, an old compact disc left to him from his deceased mother and asks her to recover its contents. Will Cat help him find a working CD reader? And what will they discover among the contents of the disc?

Watch at http://youtu.be/SPF-xzMarlg

Multimedia resources

Team Digital Preservation

https://www.youtube.com/user/wepreserve

Snow Byte & the Seven Formats

http://youtu.be/TfMgOKy9bPw